

On Posterior Consistency of Tail Index for Bayesian Kernel Mixture Models

Cheng Li ^{*1}, Lizhen Lin ^{†2}, and David B. Dunson ^{‡1}

¹Department of Statistical Science, Duke University

²Department of Statistics and Data Science, The University of Texas at Austin

Abstract

Asymptotic theory of tail index estimation has been studied extensively in the frequentist literature on extreme values, but rarely in the Bayesian context. We investigate whether popular Bayesian kernel mixture models are able to support heavy tailed distributions and consistently estimate the tail index. We show that posterior inconsistency in tail index is surprisingly common for both parametric and nonparametric mixture models. We then present a set of sufficient conditions under which posterior consistency in tail index can be achieved, and verify these conditions for Pareto mixture models under general mixing priors.

Key words: heavy tailed distribution, kernel mixture model, normalized random measures, posterior consistency, tail index.

1 Introduction

Datasets from a variety of fields, such as environmental science, finance, industrial engineering, and telecommunications, demonstrate heavy tailed behavior that can substantially influence statistical inference and decision making. It is of interest to develop estimation methods that can capture both the bulk of the data and the tails accurately. Bayesian kernel mixture models provide a flexible framework for density estimation with strong large sample guarantees. Some of the most popular models include finite mixtures (MFM, [39], [24]), Dirichlet process mixtures (DPM, [16], [32], [33], [14], [35]), and normalized random measure mixtures (NRM, [38], [30], [1], [15]). However, most of the existing literature on Bayesian asymptotics for density estimation, including results on posterior consistency and convergence rates, assume that the true density has either a compact support or exponentially decaying tails ([22], [23], [21], [31], [41]). Some of the posterior consistency theory applies to heavy tailed densities, but with respect to

^{*}cl332@stat.duke.edu

[†]lizhen.lin@austin.utexas.edu

[‡]dunson@duke.edu

a topology induced by distances, such as L_1 or Hellinger ([43], [47], etc.), which are insensitive to the tail behavior. There exist fundamental limitations and barriers in understanding the tail behavior of kernel mixture models and their large sample properties, especially for models with nonparametric mixing priors.

The current paper investigates theory on the tail behavior of popular Bayesian kernel mixture models, assessing whether they are suitable for modeling heavy tailed distributions. We focus on studying the tails of univariate continuous densities and assume that the true density has polynomially decaying tails. Such power law behavior has been observed in many real data applications (see [6] for a review). Denote $f_0(x)$ and $F_0(x)$ as the true density function and the true cumulative distribution function (cdf) with $x \in \mathbb{R}$. Let $\overline{F}(x) = 1 - F(x)$ be the survival function of cdf $F(x)$. For sufficiently large x , assume that

$$\overline{F}(x) = x^{-\alpha_+(F)} L_F(x), \quad (1)$$

where $\alpha_+(F) > 0$ is the *right tail index*, and $L_F(x)$ is a *slowly varying function* that satisfies $\lim_{y \rightarrow +\infty} L_F(xy)/L_F(y) = 1$ for any $x > 0$. The decay rate in the right tail of $f_0(x)$ can be characterized by the right tail index $\alpha_+(F_0)$, up to some slowly varying function $L_{F_0}(x)$. The left tail index can be defined similarly. If $\alpha_+(F_0) \in (0, +\infty)$, then from extreme value theory, the distribution $F_0(x)$ falls within the class of *Fréchet maximum domain of attraction* (FMDA, [2]). Examples of cdfs satisfying (1) include the Pareto distribution, the Student's t distribution, and the F distribution.

We study theoretical properties of the posterior distribution of the right tail index $\alpha_+(F)$ in a Bayesian framework. In particular, we consider the kernel mixture model:

$$f(x) = \int k(x; \boldsymbol{\theta}) dG(\boldsymbol{\theta}), \quad G \sim \pi(G; \boldsymbol{\xi}), \quad (2)$$

where $k(\cdot; \boldsymbol{\theta})$ is a univariate kernel function with parameter $\boldsymbol{\theta}$ such that $\int k(x; \boldsymbol{\theta}) dx = 1$ for all $\boldsymbol{\theta}$, G is a mixing measure of $\boldsymbol{\theta}$, and π is the prior on G with hyperparameters $\boldsymbol{\xi}$. This model is quite general, covering the aforementioned MFM, DPM and NRM models as special cases.

We will answer two critical questions for understanding how Model (2) can handle heavy tailed densities: (i) what choices of kernels and priors for the mixing measure can generate density functions with *tail indices varying* in a reasonable range, and (ii) under what types of conditions can one guarantee that the tail indices from the posterior are close to the tail index of the true density. The first question is related to whether the Bayesian kernel mixture model is capable of flexibly fitting heavy tailed distributions with different decay rates. The second question is on the frequentist asymptotic properties of Bayesian models estimating the tail index, requiring substantial extension of the scope of existing theory for Bayesian density estimation.

There is a rich literature on frequentist estimation of the tail index. Most of the estimators are constructed from tail order statistics, such as the Hill's estimator ([29], [8]), the Pickands' estimator ([37]) and their variations. The Hill's estimator is consistent ([34]) and asymptotically normal with appropriate choices of the tail order statistics for certain nonparametric classes of

distributions ([26], [25]). Minimax rates for the tail index have been obtained under different classes of distributions ([27], [12], [13], [36], [5]), and they are attainable through adaptive estimators ([28], [5], [4]).

However, there is a lack of understanding of the properties of likelihood-based approaches. The limited Bayesian literature has focused on a peak-over-threshold (POT) strategy, with the tail of the density over a high threshold t assumed to follow a generalized Pareto distribution. If $F(x)$ belongs to FDMA with right tail index $\alpha_+(F)$, then as the threshold t becomes large, the right excess distribution $\tilde{F}(x) = F(t+x)/F(t)$ for $x > 0$ converges in law to a generalized Pareto distribution with tail index $\alpha_+(F)$. Posterior sampling schemes have been discussed in, for example, [17], [3], [42], [9], [10], [45], [20]. The POT strategy can be viewed as artificial in choosing different models below and above the threshold, with the restriction of a parametric tail. The kernel mixture model allows one to choose a single flexible model for all of the data including the tails. [44] argue in favor of such an approach in using DPMs of Pareto kernels.

The rest of the paper is organized as follows. In Section 2 we show that in general, location-scale kernel mixture models cannot generate densities with varying tail indices, even if the kernel is heavy tailed. In particular, our results reveal that in many cases, the posterior distribution under the mixture model can only generate distributions with a *singleton index*. In Section 3, we provide general sufficient conditions for Bayesian posterior consistency of tail index. These conditions are then verified for the example of Pareto kernel mixtures. Section 4 concludes with discussions. All technical proofs are included in the appendix.

2 Inconsistency of Tail Index with Fixed Tailed Kernels

2.1 Posterior Consistency for Tail Index

For distribution function $F(x)$ with density $f(x)$, for $x \in \mathbb{R}$, the right and left tail indices are defined as

$$\begin{aligned}\alpha_+(F) &= \lim_{x \rightarrow +\infty} \frac{-\log \bar{F}(x)}{\log x} = \lim_{x \rightarrow +\infty} \frac{-\log P_F(X > x)}{\log x}, \\ \alpha_-(F) &= \lim_{x \rightarrow -\infty} \frac{-\log F(x)}{\log(-x)} = \lim_{x \rightarrow -\infty} \frac{-\log P_F(X \leq x)}{\log(-x)},\end{aligned}\tag{3}$$

where $P_F(\cdot)$ denotes the probability evaluated under the cdf $F(x)$. In the following, we will mainly discuss properties related to $\alpha_+(F)$ and all the arguments can apply similarly to $\alpha_-(F)$. The limits in (3) are well defined because $F(x)$ and $\bar{F}(x)$ are monotone functions. Both $\alpha_+(F)$ and $\alpha_-(F)$ take values in $[0, +\infty]$. For the right tail, $\alpha_+(F) = +\infty$ represents a thin tailed cdf, such as the exponential distribution, the normal distribution, the log-normal distribution; $\alpha_+(F) = 0$ represents a super heavy tailed cdf, such as the log-Pareto distribution ([7]). The cases that are of most interest are when $\alpha_+(F) \in (0, +\infty)$.

Let $\Pi(df)$ be the prior distribution of the density function $f \in \mathcal{F}$ (with corresponding cdf $F \in \mathcal{F}$), and let $\mathbf{X}^n = \{X_1, \dots, X_n\}$ be a sample of i.i.d data from true distribution F_0 with

density f_0 . The posterior distribution of $\Pi(\cdot|\mathbf{X}^n)$ evaluated at some measurable set $A \subseteq \mathcal{F}$ is given by

$$\Pi(A|\mathbf{X}^n) = \frac{\int_A \prod_{i=1}^n f(X_i) \Pi(df)}{\int_{\mathcal{F}} \prod_{i=1}^n f(X_i) \Pi(df)}. \quad (4)$$

Because the true tail index $\alpha_{0+} = \alpha_+(F_0)$ is unknown *a priori*, we hope that as the sample size n increases, a density $F(x)$ generated by the posterior $\Pi(\cdot|\mathbf{X}^n)$ will have a tail index $\alpha_+(F)$ sufficiently close to the truth α_{0+} . This is formalized in the following definition.

Definition 1. For any distribution $F \in \mathcal{F}$ and $\epsilon > 0$, the ϵ -(right) tail index neighborhood of F is

$$B_{\alpha_+}(F, \epsilon) \equiv \{H \in \mathcal{F} : |\alpha_+(H) - \alpha_+(F)| < \epsilon\}.$$

Definition 2. The posterior distribution $\Pi(\cdot|\mathbf{X}^n)$ is consistent for the (right) tail index if for any $\epsilon > 0$, as $n \rightarrow \infty$,

$$\Pi(B_{\alpha_+}^c(F_0, \epsilon)|\mathbf{X}^n) \rightarrow 0, \text{ in } P_{F_0}^{(n)} \text{ probability.}$$

Definition 2 is similar to the usual definition of posterior consistency for density estimation, but uses the tail index neighborhood in Definition 1. It requires that the posterior probability assigns zero mass to distributions outside ϵ -balls of F_0 as the sample size goes to infinity. The difference between the tail indices of two distributions used in Definition 1 is a pseudometric, since different distributions can have the same tail index. In general, this pseudometric has no relation with metrics used for weak or strong topology, such as the L_1 and Hellinger metrics. Posterior consistency based on L_1 or Hellinger provides a different way of measuring the concentration of the posterior compared to posterior consistency of tail index. As illustration, we take the definition of strong posterior consistency under L_1 metric as an example. If the posterior is strongly consistent at the true density f_0 with cdf F_0 , then for any $\epsilon > 0$, as $n \rightarrow \infty$,

$$\Pi(F \in \mathcal{F} : d_{TV}(F, F_0) \geq \epsilon | \mathbf{X}^n) \rightarrow 0,$$

in $P_{F_0}^{(n)}$ probability, where d_{TV} is the total variation metric equivalent to L_1 metric. The only tail information we can obtain from this relation is that for any $x > 0$,

$$\Pi(F \in \mathcal{F} : |\overline{F}(x) - \overline{F}_0(x)| \geq \epsilon | \mathbf{X}^n) \rightarrow 0, \quad (5)$$

in $P_{F_0}^{(n)}$ probability. However, this relation is fundamentally different from the posterior consistency in Definition 2, which can be rewritten from (3) as

$$\Pi\left(F \in \mathcal{F} : \lim_{x \rightarrow +\infty} \left| \frac{\log(\overline{F}(x)/\overline{F}_0(x))}{\log x} \right| \geq \epsilon | \mathbf{X}^n\right) \rightarrow 0, \quad (6)$$

in $P_{F_0}^{(n)}$ probability. In general, neither condition (5) nor (6) implies the other. Hence, although posterior consistency, such as weak and strong consistency of density estimation, is already well known for models like (7) below (see for example [22], [43], [47]), posterior consistency of tail index does not directly follow and requires further study.

We define some useful notations. For two functions $g_1(x)$ and $g_2(x)$ with $x \in \mathbb{R}^+$, $g_1(x) \prec g_2(x)$ and $g_2(x) \succ g_1(x)$ denote the relation $\lim_{x \rightarrow +\infty} g_1(x)/g_2(x) = 0$; $g_1(x) \preceq g_2(x)$ and $g_2(x) \succeq g_1(x)$ denote the relation $\limsup_{x \rightarrow +\infty} g_1(x)/g_2(x) < \infty$. P_F and E_F represent the probability and the expectation under the cdf $F(x)$. A random variable X has the decomposition $X = X_+ - X_-$, where $X_+ = \max(X, 0)$ and $X_- = \max(-X, 0)$.

2.2 Inconsistency in Location-Scale Mixture Models

In this section, we focus on a special case of Model (2), the location-scale mixture model

$$f(x) = \int \frac{1}{\sigma} k\left(\frac{x - \mu}{\sigma}\right) dG(\mu, \sigma), \quad G \sim \pi(G; \boldsymbol{\xi}), \quad (7)$$

where $k(\cdot)$ is a kernel density function and the parameter $\boldsymbol{\theta} = (\mu, \sigma)$ consists of the location parameter μ and the scale parameter σ . We assume that the kernel $k(\cdot)$ has full support on \mathbb{R} . Frequentist asymptotic properties of this model have been extensively studied in the Bayesian nonparametrics literature. Both weak and strong posterior consistency of Model (7) have been discussed in [22], [43], [47], etc. Theorem 3.3 of [43] established weak consistency of Model (7) when the true density $f_0(x)$ has a very thick polynomially decaying tail, with the tail index in $(0, 1)$. However, in the following, we will show that weak consistency, and even strong consistency based on L_1 or Hellinger distance, is insufficient for meaningful Bayesian inference of the tail index. Surprisingly, for many commonly used priors $\pi(G; \boldsymbol{\xi})$, the tail index of $f(x)$ generated from Model (7) can only take *one single value*, implying that the posterior is inconsistent in the tail index unless we know the true α_{0+} *a priori*.

For the MFM model ([39], [24]), $f(x)$ in Model (7) is specified as a finite mixture of N components ($N \in \mathbb{Z}^+$), and a further prior distribution is imposed on N . In more details, the model is given as,

$$\begin{aligned} f(x) &= \sum_{i=1}^N \frac{w_i}{\sigma_i} k\left(\frac{x - \mu_i}{\sigma_i}\right), \\ (\mu_i, \sigma_i)_{i=1}^N \Big| N &\stackrel{\text{iid}}{\sim} G_0(\mu, \sigma) \\ (w_1, \dots, w_N) \Big| N &\sim \text{Dirichlet}(a, \dots, a), \text{ for some } a > 0 \\ N &\sim \pi(N) \text{ for } N = 1, 2, \dots, \end{aligned} \quad (8)$$

The following theorem characterizes the tail index of $F(x)$ generated by Model (8).

Theorem 1. *Suppose that $G_0(\mu, \sigma)$ is a continuous distribution for (μ, σ) . Then for any distribution $F(x)$ with density $f(x)$ drawn from Model (8), the range of $\alpha_+(F)$ is almost surely a singleton. In other words, almost surely all $F(x)$ drawn from the MFM model have the same tail index.*

In the finite mixture model given in (8), the tail indices of different $F(x)$ are all the same, since all $F(x)$ are finite mixtures and their tail indices are solely determined by the tail heaviness

of the kernel $k(\cdot)$. A heavy tailed kernel will only make the tails of $F(x)$ heavy, but not be able to generate varying tail heaviness. This limitation immediately leads to posterior inconsistency of tail index.

We now investigate the more complicated example where $G(\mu, \sigma)$ has a nonparametric NRM prior. In the theorems to follow, we adopt similar NRM notations as those in [1] and [15]. Let \mathbb{X} be a complete and separable metric space endowed with the Borel σ -algebra. We consider a completely random measure $\tilde{G}(x) = \sum_{i \geq 1} \tilde{J}_i \delta_{X_i}(x)$ for $x \in \mathbb{X}$ such that $\{X_i\}_{i \geq 1}$ and nonnegative $\{\tilde{J}_i\}_{i \geq 1}$ are independent sequences of random variables, ignoring jumps at nonrandom positions. The joint distribution of $\{\tilde{J}_i\}_{i \geq 1}$ and $\{X_i\}_{i \geq 1}$ is characterized by the Lévy intensity $\nu(dv, dx)$ through the Laplace transformation of $\tilde{G}(x)$ (for $s > 0$):

$$E \left[e^{-s\tilde{G}(A)} \right] = \exp \left\{ - \int_{\mathbb{R}^+ \times A} (1 - e^{-sv}) \nu(dv, dx) \right\}.$$

We consider the homogenous NRM where the Lévy intensity can be factorized as $\nu(dv, dx) = \rho(dv)G_0(dx)$. $\rho(dv)$ is the Lévy intensity for the nonnegative masses $\{\tilde{J}_i\}_{i \geq 1}$, and $\{X_i\}_{i \geq 1}$ are independent draws from the nonatomic probability measure G_0 , also called the “base measure”. Then a NRM is defined as $G(x) = \sum_{i \geq 1} J_i \delta_{X_i}(x)$ with $J_i = \tilde{J}_i / \sum_{i \geq 1} \tilde{J}_i$. Equivalently, the cdf $G(x)$ also has the representation $G(x) = S(G_0(x))/S(1)$, where $\{S(t), t \geq 0\}$ is a subordinator with Lévy intensity measure $\rho(dv)$ (see for example [38]). For all the theorems in this section, we assume that $\rho(dv)$ satisfies $\int_0^\infty \rho(dv) = +\infty$ and $\int_0^\infty (1 - e^{-v})\rho(dv) < +\infty$ such that $0 < \sum_{i \geq 1} \tilde{J}_i < \infty$ and $0 < S(1) < +\infty$ almost surely.

The following proposition is a combination of Theorem 1 in [18] and Lemmas 4 and 5 in [19], which can be used as a fundamental tool in studying the tail behavior of a NRM.

Proposition 1. (*Fristedt [18], Fristedt and Pruitt [19]*) Suppose $\{S(t), t \geq 0\}$ is a subordinator with Lévy intensity measure $\rho(dv)$ for $v \in \mathbb{R}^+$. Define the Laplace exponent $\Psi(s)$ for $S(t)$ as $\Psi(s) = \int_0^{+\infty} (1 - e^{-sv})\rho(dv)$ and let Ψ^{-1} be the inverse function of Ψ . Define the functionals $R_L(h) = \liminf_{t \rightarrow 0^+} S(t)/h(t)$ and $R_U(h) = \limsup_{t \rightarrow 0^+} S(t)/h(t)$.

(i) For $\gamma > 0$, let

$$h_\gamma(x) = \frac{\log |\log x|}{\Psi^{-1} \left(\frac{\gamma \log |\log x|}{x} \right)}. \quad (9)$$

Then

$$\begin{aligned} R_L(h_\gamma) &\leq \gamma \text{ a.s. if } \gamma < 1, \\ R_L(h_\gamma) &\geq \gamma - 1 \text{ a.s. if } \gamma > 1. \end{aligned}$$

(ii) If $h(x)$ is locally convex in $x \in [0, \epsilon]$ for some small $\epsilon > 0$, then

$$\begin{aligned} R_U(h) &= 0 \text{ a.s. if } \int_0^\epsilon \rho[h(x), +\infty) dx < +\infty, \\ R_U(h) &= +\infty \text{ a.s. if } \int_0^\epsilon \rho[h(x), +\infty) dx = +\infty. \end{aligned}$$

Building on the proposition, we obtain the following theorem on the tail index.

Theorem 2. *Suppose $G(x)$ is a homogeneous NRM with the Lévy intensity measure $\rho(dv)G_0(dx)$ for $v \in \mathbb{R}^+, x \in \mathbb{R}$ where G_0 is a continuous probability measure on \mathbb{R} . Then*

- (i) $\alpha_+(G) = 0$ a.s., if there exists a function $h_\gamma(x)$ defined as in (9) with $\gamma > 1$, such that $\lim_{x \rightarrow +\infty} -\log h_\gamma(\overline{G}_0(x))/\log x = 0$.
- (ii) $\alpha_+(G) = +\infty$ a.s., if there exists a function $h(x)$ such that
 - (a) $h(x)$ is locally convex in $x \in [0, \epsilon)$ for some small $\epsilon > 0$;
 - (b) $\int_0^\epsilon \rho[h(x), +\infty)dx < +\infty$;
 - (c) $\lim_{x \rightarrow +\infty} -\log h(\overline{G}_0(x))/\log x = +\infty$.

The results from [18] and [19] indicate that in general it is not possible to find a function $h(x)$ that matches perfectly with the limiting behavior of a subordinator at $x = 0$. In other words, the limsup and the liminf in Proposition 1 do have gaps between them in most cases. Proper choices of the functions $h_\gamma(x)$ in (i) and $h(x)$ in (ii) will lead to sharp lower and upper bounds for the tail index of a NRM $G(x)$. The next theorem describes how these bounds for a NRM $G(x)$ can be used to characterize the tail behavior of a mixture density drawn from Model (7).

Theorem 3. *Suppose in Model (7), $G(\mu, \sigma)$ is a homogeneous NRM with Lévy intensity measure $\rho(dv)G_0(dv, d\sigma)$ for $v \in \mathbb{R}^+$ and $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$, where $G_0(\mu, \sigma)$ is a continuous probability measure on $\mathbb{R} \times \mathbb{R}^+$. Let $G_{0,\mu}(\mu)$ and $G_{0,\sigma}(\sigma)$ be the marginal measures of $G_0(\mu, \sigma)$. Assume that $G_{0,\mu}$ is symmetric about zero. If both $G_{0,\mu}(\mu)$ and $G_{0,\sigma}(\sigma)$ satisfy either (i) or (ii) in Theorem 2, then for any distribution $F(x)$ with density $f(x)$ sampled from Model (7), the range of $\alpha_+(F)$ is almost surely a singleton.*

Theorem 3 indicates that the tail indices of all distributions $F(x)$ drawn from Model 7 are almost surely the same, if each of the two marginals $G_{0,\mu}(\mu)$ and $G_{0,\sigma}(\sigma)$ satisfies either (i) or (ii) in Theorem 2. Again this leads to posterior inconsistency of tail index, by similar arguments after Theorem 1. Theorem 3 and its proof also lead to two other interesting implications. First, if the conditions for the two marginals of the base measure hold, then the tail index of $F(x)$ only depends on the tail indices of the two marginals, but not on the full joint base measure $G_0(\mu, \sigma)$. Second, whether $\alpha_+(F)$ is the same for all F does not depend on the tail behavior of the kernel $k(\cdot)$, even if $k(\cdot)$ is a heavy tailed kernel.

Remark 1. The assumption of symmetric $G_{0,\mu}$ is only used as a sufficient condition for the case where $G_{0,\mu}$ satisfies (ii) of Theorem 2 and $G_{0,\sigma}$ satisfies (i) of Theorem 2, in other words, the case where G_μ has thin left and right tails and G_σ has a super heavy right tail. The assumption of symmetric $G_{0,\mu}$ is not necessary for the conclusions of Theorem 3 to hold when both G_μ and G_σ are thin tailed, and when G_μ has a super heavy right tail. Details of the proof can be found in the appendix.

We make the tail conditions on $G_{0,\mu}(\mu)$ and $G_{0,\sigma}(\sigma)$ more concrete for the special cases of DP and normalized generalized Gamma process (NGGP, [1], [15]) mixture models. It turns

out that there is a large class of measures that satisfy the condition (i) or (ii) in Theorem 3, including both thin tailed distributions and heavy tailed distributions.

Theorem 4. *Suppose in Model (7), $G(\mu, \sigma) \sim \text{DP}(a, G_0(\mu, \sigma))$ with $a > 0$ and $G_0(\mu, \sigma)$ a continuous probability measure on $\mathbb{R} \times \mathbb{R}^+$. Assume that $G_{0,\mu}$ is symmetric about zero. If both $G_{0,\mu}(\mu)$ and $G_{0,\sigma}(\sigma)$ satisfy either one of the following conditions for a cdf $G_0(x)$ (with $x > 0$):*

- (i) $\overline{G}_0(x) \asymp \log \log x / \log x$,*
- (ii) $\overline{G}_0(x) \prec 1 / [(\log x) \cdot (\log \log x)^\delta]$ for some $\delta > 1$,*

then for any distribution $F(x)$ with density $f(x)$ sampled from Model (7), the range of $\alpha_+(F)$ is almost surely a singleton.

The proof of Theorem 4 involves the tail behavior of a DP, which has been studied in [11]. Conditions (i) and (ii) correspond to conditions (i) and (ii) in Theorem 3. As a result of the theorem, most distributions $G_{0,\mu}$ and $G_{0,\sigma}$ with either heavier or thinner tails than $1/\log x$ will lead to a single value of tail index for all $F(x)$ in the DP mixture model, and therefore the posterior of tail index is inconsistent. For example, in the popular DP mixture of normals ([14]), the marginal distributions of the base measure for μ and σ^2 are the Student's t distribution and the inverse gamma distribution, both of which have much thinner tails than $1/\log x$. Therefore, the normal-inverse gamma prior for DP mixture of normals is inconsistent for the tail index. In contrast, Theorem 3.3 of [43] has shown that such a normal-inverse gamma base measure is sufficient for posterior weak consistency, even if the true density is heavy tailed with a tail index in $(0, 1)$. This implies that the conditions required for tail index consistency are more stringent than those for usual posterior consistency. We emphasize again that the kernel here plays an inconsequential role due to Theorem 3, regardless of its tail thickness.

An important implication of Theorem 4 is that the bounds in (i) and (ii) are not far from each other. As a result, not many distributions have been left out by (i) and (ii). Basically, only those measures $G_0(x)$ that decay at a similar rate to $1/\log x$ are not covered by them. In fact, the only combination that is not covered by Theorem 4 is the case where both $G_{0,\mu}$ and $G_{0,\sigma}$ decay at rates similar to $1/\log x$. When this happens, the tail index of $F(x)$ drawn from Model (7) can vary in an interval, possibly $(0, +\infty)$. Hence, almost all DP mixture models currently used in practice are inconsistent for the tail index.

The next theorem shows the inconsistency of tail index for the general NGGP mixture model, denoted by $\text{NGGP}(a, \kappa, \tau, G_0(\mu, \sigma))$. Its Lévy intensity measure is given by $\rho(dv)dG_0(\mu, \sigma) = \frac{a}{\Gamma(1-\kappa)} v^{-\kappa-1} e^{-\tau v} dv dG_0(\mu, \sigma)$, where $a > 0$, $\kappa \in [0, 1)$ and $\tau > 0$. The NGGP class includes most of the discrete random probability measures in the Bayesian nonparametric literature. For example, the class includes DP as $\text{NGGP}(a, 0, 1, G_0)$, the normalized-inverse Gaussian process as $\text{NGGP}(1, 1/2, \tau, G_0)$, and the N-stable process as $\text{NGGP}(1, \kappa, 0, G_0)$ as special cases. See [1] and [15] for discussions. For tail index inconsistency, the cases of $\kappa = 0$ (DP) and $\kappa > 0$ are different in nature, so the conclusion of Theorem 5 is also different from Theorem 4.

Theorem 5. *Suppose in Model (7), $G(\mu, \sigma) \sim \text{NGGP}(a, \kappa, \tau, G_0(\mu, \sigma))$ with $a > 0$, $\kappa \in (0, 1)$, $\tau \geq 0$ and $G_0(\mu, \sigma)$ is a continuous probability measure on $\mathbb{R} \times \mathbb{R}^+$. Assume that $G_{0,\mu}$ is*

symmetric about zero. If both $G_{0,\mu}(\mu)$ and $G_{0,\sigma}(\sigma)$ satisfy either one of the following conditions for a cdf $G_0(x)$ (with $x > 0$):

(i) $\overline{G}_0(x) \succ x^{-\delta}$ for all $\delta > 0$,

(ii) $\overline{G}_0(x) \prec x^{-\delta}$ for all $\delta > 0$,

then for any distribution $F(x)$ with density $f(x)$ sampled from Model (7), the range of $\alpha_+(F)$ is almost surely a singleton.

Similar to Theorem 4, here we also provide two conditions for the tail decaying rates of $G_{0,\mu}$ and $G_{0,\sigma}$, where (i) gives heavier than polynomial tails and (ii) gives thinner than polynomial tails. The gap between G_0 that satisfy (i) or (ii) in the current theorem is now larger than that in the DP case, but the theorem still has ruled out many possibilities for posterior consistency of tail index. For example, when both $G_{0,\mu}$ and $G_{0,\sigma}$ have exponentially decaying tails, the posterior of a NGGP is inconsistent for tail index. It remains unknown how the tail indices of $F(x)$ from a NGGP mixture model behave in the posterior when at least one of $G_{0,\mu}$ and $G_{0,\sigma}$ have a polynomially decaying tail.

3 Sufficient Conditions for Tail Index Consistency

3.1 Schwartz's Theorem for Posterior Consistency

In this section, we provide a series of conditions that guarantee the posterior consistency of tail index for the most general model $f \sim \Pi$. These conditions are built on the classic Schwartz's argument [40] for posterior consistency, and therefore they are simple and intuitive. We will then demonstrate the application of these sufficient conditions on Model (2) using the Pareto kernel in Section 3.3.

The Schwartz consistency theorem relies on two key conditions: the Kullback-Leibler (KL) support of the prior, and the existence of a uniformly consistent test. For two distributions F_1 and F_2 (with densities f_1 and f_2), let the KL divergence between F_1 and F_2 be $KL(F_1, F_2) \equiv E_{F_1} \log(f_1/f_2)$. Define the ϵ -KL neighborhood of the true distribution F_0 as $\mathcal{K}(F_0, \epsilon) \equiv \{F \in \mathcal{F} : KL(F_0, F) < \epsilon\}$. The KL support of the prior is stated as follows:

(KL) The true distribution F_0 is in the KL support of Π , if for any $\epsilon > 0$,
 $\liminf_{n \rightarrow \infty} \Pi(\mathcal{K}(F_0, \epsilon)) > 0$.

We do not rule out the possibility that the prior Π depends on the sample size n , since this can be conveniently incorporated into the standard posterior consistency argument (see Section 5 of [22]). It is well known that the KL support of Π implies weak consistency. Therefore (KL) is a very basic requirement for useful Bayesian models.

The other condition required in the Schwartz consistency theorem is the existence of uniformly consistent tests. For our purpose, we need a test for tail index that is able to separate F_0 from all the distributions *outside a tail index neighborhood of F_0* . The sieve \mathcal{F}_n helps when

Π has a non-compact support and the uniform test can be found on a sufficiently large compact support.

(UT) Uniform testing condition: There exists a test $\Phi_n \equiv \Phi_n(X_1, \dots, X_n)$ and a sieve \mathcal{F}_n such that

- (i) $\Pi(\mathcal{F}_n^c) \leq e^{-bn}$ for some constant $b > 0$;
- (ii) For any $\epsilon > 0$, as $n \rightarrow \infty$,

$$E_{F_0} \Phi_n \rightarrow 0, \quad \sup_{F \in B_{\alpha+}^c(F_0, \epsilon) \cap \mathcal{F}_n} E_F(1 - \Phi_n) \rightarrow 0. \quad (10)$$

Based on a similar argument as Schwartz's consistency theorem, one can show posterior consistency of tail index under the conditions (KL) and (UT).

Theorem 6. *If both (KL) and (UT) hold true, then the posterior distribution $\Pi(\cdot | \mathbf{X}^n)$ is consistent for the (right) tail index.*

The proof follows the same thread as the usual weak consistency (see for example [22]) and is therefore omitted. Note that the uniform test in (UT) can be made exponentially fast by an argument using the Hoeffding's inequality (Theorem 2 of [22]). However, a key unanswered question is whether such a uniformly consistent test Φ_n for tail index exists. One cannot directly apply the usual asymptotic theory because Φ_n depends on our new tail index neighborhood defined through a pseudometric, whose induced topology is different from the weak topology. We instead proceed in a constructive way and pursue sufficient conditions for (UT) to hold.

3.2 Existence of Tests

In the representation (1) for a generic cdf F , let $h_F(x) = xL'_F(x)/L_F(x)$ and hence $L_F(x) = L_F(x_0) \exp\left(\int_{x_0}^x \frac{h_F(t)}{t} dt\right)$ for some fixed x_0 . Alternatively, $h_F(x)$ can be written as

$$h_F(x) = \alpha_+(F) - \frac{xf(x)}{\overline{F}(x)}.$$

For any given $F(x)$ from FMDA, the von-Mises theorem (see e.g. Proposition 2.1 of [2]) says that

$$\lim_{x \rightarrow +\infty} \frac{xf(x)}{\overline{F}(x)} = \alpha_+(F),$$

i.e. $\lim_{x \rightarrow +\infty} h_F(x) = 0$. Bounding the magnitude of $h_F(x)$ is crucial in showing the existence of uniform tests for $\alpha_+(F)$. In the Bayesian framework, $h_F(x)$ with $F \sim \Pi$ needs to be controlled in a uniform way on a sieve with large prior probability. In light of this, we have the following theorem on the existence of tests.

Theorem 7. *Suppose the following conditions hold:*

- (i) *There exist finite constants $x_0 \geq 1$ and $c_0 > 0$, such that $e^{-c_0} \leq L_F(x_0) \leq e^{c_0}$ for all $F \in \mathcal{F}_{1n}$ where \mathcal{F}_{1n} is a sieve satisfying $\Pi(\mathcal{F}_{1n}^c) < e^{-c_1 n}$ for some constant $c_1 > 0$;*
- (ii) *There exists an envelope function $\overline{h}_n(x) = B_n x^{-\tau_n}$ for some positive n -dependent sequences*

B_n and τ_n , such that $|h_F(x)| \leq \bar{h}_n(x)$ for all $F \in \mathcal{F}_{2n}$ and all $x \geq x_0$, where \mathcal{F}_{2n} is a sieve satisfying $\Pi(\mathcal{F}_{2n}^c) < e^{-c_2 n}$ for some constant $c_2 > 0$;

(iii) The prior Π satisfies $\Pi(\mathcal{F}_{3n}^c) < e^{-c_3 n}$ for some constant $c_3 > 0$ for $\mathcal{F}_{3n} = \{F \in \mathcal{F} : \alpha_+(F) \leq \bar{\alpha}_n\}$ and some sequence $1 \prec \bar{\alpha}_n \prec \log n$, for all sufficiently large n ;

(iv) There exists a nonnegative sequence s_n such that $1 \prec s_n \prec \bar{\alpha}_n^{-1} \log n$, $B_n \prec \min(e^{\tau_n s_n}, \tau_n \log n)$, as $n \rightarrow \infty$;

then (UT) holds.

The proof of Theorem 7 uses a recently proposed tail index estimator in [5] defined as

$$\hat{\alpha}_{s_n} = \log(\hat{p}_{s_n}) - \log(\hat{p}_{s_n+1}), \quad (11)$$

where $\hat{p}_{s_n} = n^{-1} \sum_{i=1}^n I(X_i > e^{s_n})$ and s_n is the same sequence in Condition (iv). [5] has shown that $\hat{\alpha}_{s_n}$ is a consistent estimator of $\alpha_+(F)$ for F from various classes of distributions. Therefore, a test for $H_0 : \alpha_+(F) = \alpha_{0+}$ can be $\Phi_n = I(|\hat{\alpha}_{s_n} - \alpha_{0+}| > \epsilon)$ given some $\epsilon > 0$. For our purposes, it is easier to work with $\hat{\alpha}_{s_n}$ than the Hill's estimator.

Conditions (i)-(iv) are sufficient for the existence of such tests. Among them, (i) and (ii) are mainly intended to control the slowly varying function $L_F(x)$, where we allow exceptions on sets with exponentially small prior probabilities. The choice of $x_0 \geq 1$ is mainly for convenience since $\log x > 0$ for all $x \geq x_0$. Alternatively, one can replace it with any finite $x_0 \in \mathbb{R}$ and modify the definition of logarithm function with a shift accordingly. In (ii) we specify the envelope function $\bar{h}_n(x)$ to be polynomially decaying and allow it to depend on n , which will be made clear in our later example. In the frequentist tail index literature, such control over the exponent in a slowly varying function has appeared in [12] and [13] for showing minimax rates in certain classes of distributions. The polynomially decaying $\bar{h}_n(x)$ is not restrictive because we allow $B_n \rightarrow \infty$ and $\tau_n \rightarrow 0$ as $n \rightarrow \infty$. Condition (iii) restricts the largest possible tail index on a large sieve, but the sieve will eventually cover the true F_0 as the sample size n increases. Condition (iv) determines the choice of s_n in $\hat{\alpha}_{s_n}$, which is usually a slowly increasing sequence and also interplays with the magnitude of B_n and τ_n in (ii). For posterior consistency, we only require the existence of such sequences B_n, τ_n, s_n . [5] has given the explicit choice of s_n (as well as a data-dependent version) such that $\hat{\alpha}_{s_n}$ converges at a minimax rate to $\alpha_+(F)$ for a certain class of distributions (adaptively). Conditions (i)-(iv) will be verified for Pareto mixtures in Section 3.3.

The following theorem is a consequence of Theorem 6 and Theorem 7.

Theorem 8. (*Posterior Consistency of Tail Index*) Under all assumptions of Theorem 7 and (KL), the posterior distribution $\Pi(\cdot | \mathbf{X}^n)$ of Model (2) is consistent for the (right) tail index.

The tail index inconsistency for the location-scale Model (7) in Theorem 1 and Theorem 3 is mainly caused by the fact that all densities $f(x)$ drawn from the prior Π have the same tail index. This limitation can be viewed as a lack of *tail index support* from the prior Π , since posterior consistency requires that the prior Π should put positive probability mass on an open

neighborhood $B_{\alpha+}(\epsilon)$ of F_0 for any $\epsilon > 0$. In other words, we need the following tail index support condition to hold true.

(TI) The true distribution $F_0(x)$ is in the tail index support of the prior Π , i.e. for any small $\epsilon > 0$, $\liminf_{n \rightarrow \infty} \Pi(B_{\alpha+}(F_0, \epsilon)) > 0$.

(TI) is clearly a *necessary* condition for the consistency of tail index. The failure of (TI) will cause posterior inconsistency, as shown in Theorems 3, 4 and 5 for Model (7). However, if the prior satisfies (KL) and all sufficient conditions in Theorem 7, then (TI) is also satisfied.

3.3 Example of Consistency: Mixtures of Paretos

The failure of tail index consistency in Section 2 is partly due to the structure of the location-scale mixture model (7), in which we have no control over how the mixing measure $G(\mu, \sigma)$ affects the tail index of the mixture density $f(x)$. A possible remedy is to introduce an explicit tail index parameter which is also mixed by G . An example of this type is the DPM of Paretos used in [44]. In this section, we study the mixture of simple Pareto distributions with kernel density $k(x; \alpha) = \alpha x^{-(\alpha+1)}$ whose support is $[1, +\infty)$. Because the mixture distribution from both MFM and NRM takes the form $\bar{F}(x) = \sum_{i=1}^{\infty} w_i x^{-\alpha_i}$, the right tail index is $\alpha_+(F) = \inf(\alpha_1, \alpha_2, \dots)$. To make this tail index more explicit, in the following Bayesian model, we are going to first pick α_1 as the tail index of $F(x)$ together with its weight w_1 , and then draw the other α_i and their weights w_i ($i = 2, 3, \dots$) as either a finite mixture (in MFM) or a NRM conditional on α_1 and w_1 . In this way we can guarantee that $\alpha_i > \alpha_1$ for all $i \geq 2$ such that we can conveniently control the behavior of $\alpha_+(F)$ through α_1 . The model is specified as follows.

$$\begin{aligned} f(x) \Big|_{\alpha_1, w_1, H} &= w_1 k(x; \alpha_1) + (1 - w_1) \int k(x; \alpha) dH(\alpha), \\ \alpha_1 &\sim G_1(\alpha), \text{ supp}(G_1) = (0, \bar{\alpha}_n], \\ w_1 &\sim \pi_w(w), \text{ supp}(\pi_w) = [\underline{w}_n, 1], \\ H &\Big|_{\alpha_1, w_1, H_2} \sim \Pi(H; H_2(\cdot; \alpha_1, \boldsymbol{\xi})), \\ H_2(\alpha; \alpha_1, \boldsymbol{\xi}) &= H_0(\alpha - (\alpha_1 + \tau_n); \boldsymbol{\xi}), \text{ supp}(H_0) = (0, \bar{\alpha}_n]. \end{aligned} \tag{12}$$

The notation “supp” stands for the support of a distribution. The density $f(x)$ has two mixing components. The first component $w_1 k(x; \alpha_1)$ explicitly controls the tail index of $f(x)$, and the second component is a general mixture of Paretos. α_1 in the first component determines $\alpha_+(F)$, and is drawn from G_1 whose support is bounded above by $\bar{\alpha}_n$, in order to meet Condition (iii) of Theorem 7. The second component involves a mixing probability measure H , which is drawn from the prior Π as either a finite mixture (making $f(x)$ a MFM) or an infinite mixture from nonparametric priors like NRM. If Π is a MFM prior, then H_2 denotes the distribution from which $\alpha_2, \dots, \alpha_N$ are drawn independently given the number of components N , and N itself has another prior Π_N . If Π is a homogeneous NRM prior, then H_2 denotes the base measure from which H is drawn. In both cases, H_2 is a distribution with left endpoint $\alpha_1 + \tau_n$, where

α_1 is drawn from G_1 , and τ_n is a deterministic sequence that goes to zero as $n \rightarrow \infty$. We can further model H_2 as a $(\alpha_1 + \tau_n)$ right-shifted version of a base distribution H_0 . ξ contains all the other hyperparameters of H_0 , such as the parameter to control the number of components N in MFM, or the parameters a, κ, τ in NGGP. The deterministic sequences $\underline{w}_n, \bar{\alpha}_n, \tau_n$ introduced here are mainly designed to separate the leading component $w_1 k(x; \alpha_1)$ from the other higher order mixing components, such that the sufficient conditions in Theorem 7 are satisfied. In fact, the way of isolating the leading Pareto component in Model (12) is similar to some well studied nonparametric classes of distributions in the frequentist tail index literature, such as the Hall and Welsh class ([28], [5], [4]) that satisfies $|\bar{F}(x) - Cx^{-\alpha}| \leq C'x^{-(1+\beta)}$ for $\alpha, \beta, C, C' > 0$.

A function $g(x)$ on the interval $x \in I$ is called completely monotone if the m th derivative of $g(x)$ satisfies $(-1)^m g^{(m)}(x) \geq 0$ for all $m \in \mathbb{Z}^+$. Let

$$\begin{aligned} \mathcal{CM}_e &= \{F(x) : \text{supp}(F) = [1, +\infty), \bar{F}_0(e^t) \text{ is completely monotone on } t \in [0, +\infty)\}, \\ \mathcal{P}_2 &= \left\{F(x) : \text{supp}(F) = [1, +\infty), \bar{F}(x) = Cx^{-\alpha} + O(x^{-(1+\beta)\alpha}), \right. \\ &\quad \left. \text{for some constant } \alpha > 0, \beta > 0, C > 0\right\}, \end{aligned}$$

where \mathcal{P}_2 is the class of second-order Pareto distributions. Then we characterize the class of distributions described by Model (12).

Theorem 9. *If $F(x) \in \mathcal{CM}_e \cap \mathcal{P}_2$, then $F(x)$ is in the KL support of Model (12), if the prior $\Pi(H; \alpha_1, H_0, \xi)$ is either a MFM or a homogeneous NRM.*

The KL support of Model (12) is related to the class of completely monotone functions. This is not surprising because the mixtures of Paretos are related to the mixtures of exponential distributions by the transformation $x = e^t$ in the Pareto kernel $k(x; \alpha)$. The KL support of the mixtures of exponentials includes the class of completely monotone functions (Theorem 16 in [47]), by the Hausdorff-Bernstein-Widder theorem.

The following theorem imposes further conditions on $\underline{w}_n, \bar{\alpha}_n, \tau_n$ and the prior G_1, Π, H_0 , such that Model (12) achieves posterior consistency of tail index.

Theorem 10. *Suppose the following conditions hold for Model (12):*

- (i) $F_0(x) \in \mathcal{CM}_e \cap \mathcal{P}_2$;
 - (ii) $1 \prec \bar{\alpha}_n \prec \log n$, $\tau_n \prec 1$, $\underline{w}_n \prec 1$, $\underline{w}_n \tau_n \succ \bar{\alpha}_n / \log n$, $\log(\bar{\alpha}_n / \underline{w}_n) \prec \tau_n \log n / \bar{\alpha}_n$,
- then the posterior distribution $\Pi(\cdot | \mathbf{X}^n)$ of Model (12) is consistent for the tail index.*

We briefly discuss the choice of the sequences $\underline{w}_n, \bar{\alpha}_n, \tau_n$ here as well as s_n in Theorem 7 to ensure their existence. For example, the conditions are satisfied by $\bar{\alpha}_n = \log^{1/2} n$, $\tau_n = \log^{-1/4} n$, $\underline{w}_n = \log^{-1/5} n$, $s_n = \log^{1/3} n$.

Remark 2. The densities in $\mathcal{CM}_e \cap \mathcal{P}_2$ always have nonnegative mixing coefficients, since $w_1 > 0$ and $H(\alpha)$ is a nonnegative probability measure. As a result, the KL support of Model

(12) includes mixtures such as $\overline{F}(x) = \frac{1}{2x} + \frac{1}{2x^2}$, but also has excluded some other mixtures of Paretos, such as $\overline{F}(x) = \frac{2}{x} - \frac{1}{x^2}$ in which some components may have negative coefficients. To enlarge the KL support of Model (12) and allow negative mixing coefficients, the mixing measure can be characterized as a bounded signed measure $w_1\delta_{\alpha_1} + (1 - w_1)H = H_+ - H_-$, where δ_a denotes the Dirac measure at a . Similar priors to those in Model (12) can be imposed on both H_+ and H_- and they need further restrictions to guarantee that the density $f(x)$ is nonnegative. For example, if $\overline{F}(x) = \frac{2}{x} - \frac{1}{x^2}$, then $H_+ = 2\delta_1$ and $H_- = \delta_2$. According to Theorem 4.3 of [46], the Pareto kernel mixture representation by using bounded signed mixing measure includes all distributions $F(x)$ that satisfy $\sum_{i=1}^{\infty} \left| \overline{F}^{(i)}(e^t) \right| t^i / i! < +\infty$.

4 Discussion

We have explored the theory behind the posterior consistency/inconsistency of tail index for Bayesian kernel mixture models, extending the scope of the vast literature on Bayesian consistency with respect to the weak and strong topology. We have shown that examples of inconsistency are extremely common, among the location-scale mixture models with MFM, DPM and NRM priors. There are special cases where the posterior consistency remains unknown in the DPM and NRM examples when the marginal base measures of the location and scale parameters meet certain restrictions, although we conjecture that in general the posteriors are still inconsistent in the tail index.

We have also proposed a set of sufficient conditions that lead to posterior tail index consistency, and verified them in a Pareto mixture example. The simple Pareto mixture model is mainly used for illustration, as other heavy tailed kernels with an explicit tail index parameter can also be implemented in a similar manner, such as the inverse gamma kernel, the half Student's t kernel, and the F kernel, although their consistency theory involves extra technical complexity in verifying all those sufficient conditions. It is less obvious to see how models like (12) can be generalized to mixing models with two-sided kernels, since ideally one wants to estimate both the left tail index and the right tail index of a distribution, which can be possibly different. It will be an interesting topic to further study the posterior convergence rates for Model (12) when the true $F_0(x)$ comes from certain nonparametric classes such as the Hall and Welsh class, and compare them with the frequentist adaptive estimators such as [5] and [4], which achieve the minimax rates.

Acknowledgment The authors thank Professor Jayanta Ghosh for a comment on the challenges of Bayesian asymptotics for heavy-tailed densities, which served as inspiration for this work.

Appendix

A Technical Proofs for Section 2

Lemma A.1. *Let $m > 0$ be finite, and $F(x)$ be an univariate continuous cdf. For $x > 0$,*

- (i) *If $E_F X_+^m < \infty$, then $P_F(X > x) \prec x^{-m}$;*
- (ii) *If $E_F X_+^m = \infty$, then $P_F(X > x) \succ x^{-(m+\delta)}$ for any $\delta > 0$.*

Proof of Lemma A.1:

(i) Since $E_F X_+^m < \infty$, we have $\int_0^{+\infty} x^m dF(x) < \infty$ and $\lim_{x \rightarrow +\infty} \int_x^{+\infty} t^m dF(t) = 0$. Therefore

$$\lim_{x \rightarrow +\infty} x^m \bar{F}(x) = \lim_{x \rightarrow +\infty} x^m \int_x^{+\infty} dF(t) \leq \lim_{x \rightarrow +\infty} \int_x^{+\infty} t^m dF(t) = 0,$$

which implies that $\bar{F}(x) = P_F(X > x) \prec x^{-m}$ for $x > 0$.

(ii) Suppose that the conclusion does not hold. Then there exists some $\delta > 0$ such that $P_F(X_+ > x) \preceq x^{-(m+\delta)}$, which implies that

$$\lim_{x \rightarrow +\infty} x^{m+\frac{\delta}{2}} P_F(X > x) = 0.$$

On the other hand, if we integrate by parts, we have

$$\int_0^{+\infty} x^m dF(x) = - \int_0^{+\infty} x^m d\bar{F}(x) = - \lim_{x \rightarrow +\infty} x^m \bar{F}(x) + \int_0^{+\infty} m x^{m-1} \bar{F}(x) dx.$$

In the last display, the first term is zero because $x^m \bar{F}(x) \prec x^{-\delta/2} \rightarrow 0$ as $x \rightarrow +\infty$. Using $\bar{F}(x) \leq 1$, the second term can be bounded by

$$\int_0^1 m x^{m-1} \bar{F}(x) dx + \int_1^{+\infty} m x^{m-1} \bar{F}(x) dx \leq 1 + \int_1^{+\infty} m x^{-1-\delta/2} dx < \infty.$$

Therefore we get $\int_0^{+\infty} x^m dF(x) < \infty$, which contradicts the assumption that $E_F X_+^m = \infty$. The conclusion $P_F(X > x) \succ x^{-(m+\delta)}$ must follow. ■

Lemma A.2. *Let $F(x)$ be an univariate continuous cdf with right tail index $\alpha_+(F)$.*

- (i) *If $\alpha_+(F) = 0$, then $E_F X_+^m = +\infty$ for all $m > 0$;*
- (ii) *If $\alpha_+(F) \in (0, +\infty)$, then $E_F X_+^{\alpha_+(F)-\delta} < +\infty$ if $0 < \delta \leq \alpha_+(F)$, and $E_F X_+^{\alpha_+(F)+\delta} = +\infty$ for all $\delta > 0$;*
- (iii) *If $\alpha_+(F) = +\infty$, then $E_F X_+^m < +\infty$ for all $m > 0$.*

Proof of Lemma A.2:

According to the definition of $\alpha_+(F)$, if $\alpha_+(F) \in [0, +\infty)$, then we have the relation

$$x^{-\alpha_+(F)-\eta} \prec \bar{F}(x) \prec x^{-\alpha_+(F)+\eta} \tag{A.1}$$

for any $\eta > 0$.

We first rewrite Lemma A.1 in its contrapositive form: for a given $m > 0$,

- (a) If $\overline{F}(x) \succeq x^{-m}$, then $E_F X_+^m = +\infty$;
- (b) If there exists $\delta_0 > 0$ such that $\overline{F}(x) \preceq x^{-(m+\delta_0)}$, then $E_F X_+^m < +\infty$. The proof of part (ii) in Lemma A.1 indicates that this δ_0 can be replaced by any smaller positive number.

Now we show (i)-(iii).

- (i) Assume $\alpha_+(F) = 0$. Then by the definition of $\alpha_+(F)$, we have $\overline{F}(x) \succ x^{-\delta}$ for all $\delta > 0$, which is exactly the same assumption as in (a). Therefore the conclusion follows from (a).

- (i) Assume $\alpha_+(F) \in (0, +\infty)$. Then from (A.1), $\overline{F}(x) \succ x^{-\alpha_+(F)-\delta}$ for any $\delta > 0$. Therefore, if we let $\eta = \delta$ and $m = \alpha_+(F) + \delta$ in (a), then we obtain that $E_F X_+^{\alpha_+(F)+\delta} = +\infty$.

Also from (A.1), $\overline{F}(x) \prec x^{-\alpha_+(F)+\eta}$ for any $\eta > 0$. Now in part (i) of the conclusion, fix a $\delta > 0$. When $\alpha_+(F) > 0$, we can set $m = \alpha_+(F) - \delta$ and $\eta = \delta_0 = \delta/2$ (because δ_0 can be chosen small and (b) still holds). Then $\overline{F}(x) \prec x^{-\alpha_+(F)+\eta} \preceq x^{-(m+\delta)} + \delta_0 = x^{-(m+\delta_0)}$ and (b) gives $E_F X_+^{\alpha_+(F)-\delta} < +\infty$.

- (iii) Assume $\alpha_+(F) = +\infty$. Then the definition of $\alpha_+(F)$ implies that $\overline{F}(x) \prec x^{-m}$ for all $m > 0$. For a fixed m , since $\overline{F}(x) \preceq x^{-(m/2+m/2)}$, we can set $\delta_0 = m/2$ in (b) and obtain that $E_F X_+^{m/2} < +\infty$. Because m can be arbitrary, this implies that $E_F X_+^m < +\infty$ for all $m > 0$. ■

Lemma A.3. *Let $m > 0$.*

- (i) *For any $x, y \in \mathbb{R}$, there exists a constant C_m that only depends on m , such that*

$$[(x+y)_+]^m \leq C_m (x_+^m + y_+^m).$$

- (ii) *For any $x \geq 0, y \geq 0$, there exists a constant c_m that only depends on m , such that*

$$(x+y)^m \geq c_m (x^m + y^m).$$

Proof of Lemma A.3:

- (i) For $m \geq 1$, $C_m = 2^{m-1}$. For $m \in (0, 1)$, $C_m = 1$.
- (ii) Let $f(t) = t^m + (1-t)^m$ and $t \in [0, 1]$. If $m \geq 1$, then $\max_{t \in [0, 1]} f(t) = 1$ and set $c_m = 1$. If $m \in (0, 1)$, then $\max_{t \in [0, 1]} f(t) = 2^{1-m}$ and set $c_m = 2^{m-1}$. Now let $t = x/(x+y)$ and the conclusion follows. ■

Lemma A.4. *Suppose $f(x)$ is a density drawn from Model (7) with cdf $F(x)$. Let $K(\cdot)$ be the cdf of $k(\cdot)$. Then*

$$E_F X_+^m \geq c_m (E_{G_\mu} \mu_+^m \cdot \overline{K}(0) + E_K X_+^m \cdot E_{G_{\mu, \sigma}} [\sigma^m I(\mu \geq 0)]), \quad (\text{A.2})$$

$$E_F X_+^m \geq C_m^{-1} E_K X_+^m \cdot E_{G_\sigma} \sigma^m - E_{G_\mu} \mu_-^m, \quad (\text{A.3})$$

$$E_F X_+^m \leq C_m (E_{G_\mu} \mu_+^m + E_K X_+^m \cdot E_{G_\sigma} \sigma^m), \quad (\text{A.4})$$

where $\overline{K}(0) = P_K(X \geq 0)$ (the probability of $X \geq 0$ if X has the density $k(x)$), G_μ and G_σ are the marginal distributions of $G_{\mu, \sigma}$, and c_m, C_m are defined in Lemma A.3.

Proof of Lemma A.4:

Let $I(\cdot)$ be the indicator function. We have

$$\begin{aligned} E_F X_+^m &= \iint x^m I(x \geq 0) \frac{1}{\sigma} k\left(\frac{x - \mu}{\sigma}\right) dG(\mu, \sigma) dx \\ &= \iint (\mu + \sigma y)^m I(\mu + \sigma y \geq 0) k(y) dG(\mu, \sigma) dy. \end{aligned} \quad (\text{A.5})$$

Then we give lower and upper bounds for (A.5). Notice that $\sigma \geq 0$ always holds and $I(\mu + \sigma y \geq 0) \geq I(\mu \geq 0)I(y \geq 0)$. Based on (A.5) and part (ii) of Lemma A.3, we have:

$$\begin{aligned} E_F X_+^m &\geq \iint (\mu + \sigma y)^m I(\mu \geq 0) I(y \geq 0) k(y) dG(\mu, \sigma) dy \\ &\geq \iint c_m(\mu_+^m + \sigma^m y_+^m) I(\mu \geq 0) I(y \geq 0) k(y) dG(\mu, \sigma) dy \\ &= c_m (E_{G_\mu} \mu_+^m \cdot \overline{K}(0) + E_K X_+^m \cdot E_{G_{\mu, \sigma}} [\sigma^m I(\mu \geq 0)]), \end{aligned}$$

which is (A.2).

On the other hand, part (i) of Lemma A.3 implies

$$\begin{aligned} (\sigma y)_+^m &\leq C_m [(\mu + \sigma y)_+^m + (-\mu)_+^m] \\ \implies (\mu + \sigma y)_+^m &\geq C_m^{-1} \sigma^m y_+^m - \mu_-^m. \end{aligned}$$

since $(-\mu)_+ = \mu_-$. This together with (A.5) gives

$$\begin{aligned} E_F X_+^m &= \iint (\mu + \sigma y)_+^m k(y) dG(\mu, \sigma) dy \\ &\geq \iint [C_m^{-1} \sigma^m y_+^m - \mu_-^m] k(y) dG(\mu, \sigma) dy \\ &\geq C_m^{-1} E_K X_+^m \cdot E_{G_\sigma} \sigma^m - E_{G_\mu} \mu_-^m, \end{aligned}$$

which is (A.3).

By part (i) of Lemma A.3

$$\begin{aligned} E_F X_+^m &= \iint [(\mu + \sigma y)_+]^m dG(\mu, \sigma) dy \\ &\leq \iint C_m(\mu_+^m + \sigma^m y_+^m) k(y) dG(\mu, \sigma) dy = C_m (E_{G_\mu} \mu_+^m + E_K X_+^m \cdot E_{G_\sigma} \sigma^m), \end{aligned}$$

which is (A.4). ■

Proof of Theorem 1:

For Model (8), the marginal distributions G_μ and G_σ are both finite mixtures at the points $\mu_{i=1}^N$ and $\sigma_{i=1}^N$ respectively. Because $G_0(\mu, \sigma)$ is a continuous distribution, we have $0 \leq E_{G_\mu} \mu_+^m < +\infty$, $0 \leq E_{G_\mu} \mu_-^m < +\infty$ and $0 < E_{G_\sigma} \sigma^m < +\infty$ for all $m > 0$. We can use Lemma A.2 to determine the relation between $\alpha_+(F)$ and $\alpha_+(K)$. According to Lemma A.4, whether $E_F X_+^m$ is finite or not for a given m is solely determined by whether $E_K X_+^m$ is finite or not. The analysis goes as follows:

(i) If $\alpha_+(K) = +\infty$, then by Lemma A.2 $E_K X_+^m < +\infty$ for all $m > 0$. The upper bound (A.4) implies that $E_F X_+^m < +\infty$ for all $m > 0$. Hence, $\alpha_+(F) = +\infty$ by Lemma A.2.

(ii) If $\alpha_+(K) = 0$, then by Lemma A.2, $E_K X_+^m = +\infty$ for all $m > 0$. The lower bound (A.3) implies that $E_F X_+^m = +\infty$ for all $m > 0$. Then by setting $m = 0$ in (ii) of Lemma A.1 we can see that $\alpha_+(F) = 0$.

(iii) If $\alpha_+(K) \in (0, +\infty)$, then $E_K X_+^m < +\infty$ for $m < \alpha_+(K)$ and $E_K X_+^m = +\infty$ for $m > \alpha_+(K)$. Then by (A.3), $E_F X_+^m < +\infty$ for $m < \alpha_+(K)$, and by (A.4), $E_K X_+^m = +\infty$ for $m > \alpha_+(K)$. Apply Lemma A.1 and we can see that $\alpha_+(F) = \alpha_+(K)$.

In sum, $\alpha_+(F) = \alpha_+(K)$ in all three cases and thus $\alpha_+(F)$ is almost surely a singleton. \blacksquare

Proof of Theorem 2:

The proof is a direct application of Proposition 1.

(i) By the stationary increment property of subordinators, $S(1-t)$ has the same distribution as $S(1) - S(t)$ for $t \in (0, 1)$. Therefore for $\gamma > 1$ and h_γ defined in (9), part (i) of Proposition 1 implies

$$\liminf_{t \rightarrow 1^-} \frac{S(1) - S(t)}{h_\gamma(1-t)} \geq \gamma - 1 \text{ a.s.}$$

Let $t = G_0(x)$ and we have

$$\liminf_{x \rightarrow +\infty} \frac{\overline{G}(x)S(1)}{h_\gamma(\overline{G}_0(x))} \geq \gamma - 1 \text{ a.s.} \quad (\text{A.6})$$

since $G(x) = S(G_0(x))/S(1)$. Our assumptions $\rho(dv)$ satisfies $\int_0^\infty \rho(dv) = +\infty$ and $\int_0^\infty (1 - e^{-v})\rho(dv) < +\infty$ guarantee that $0 < S(1) < +\infty$. Therefore, (A.6) implies that almost surely for all such NRM $G(x)$, $\overline{G}(x) \succeq h_\gamma(\overline{G}_0(x))$, which implies that

$$\alpha_+(G) = \lim_{x \rightarrow +\infty} \frac{-\log \overline{G}(x)}{\log x} \preceq \lim_{x \rightarrow +\infty} \frac{-\log h_\gamma(\overline{G}_0(x))}{\log x} = 0.$$

Therefore $\alpha_+(G) = 0$.

(ii) For such $h(x)$ that satisfies (a)(b)(c), by similar argument as above, we apply part (ii) of Proposition 1 and obtain that

$$\limsup_{x \rightarrow +\infty} \frac{\overline{G}(x)S(1)}{h(\overline{G}_0(x))} = 0 \text{ a.s.}$$

which implies that $\overline{G}(x) \preceq h(\overline{G}_0(x))$ almost surely for all such NRM $G(x)$. Therefore

$$\alpha_+(G) = \lim_{x \rightarrow +\infty} \frac{-\log \overline{G}(x)}{\log x} \succeq \lim_{x \rightarrow +\infty} \frac{-\log h(\overline{G}_0(x))}{\log x} = +\infty,$$

which means $\alpha_+(G) = +\infty$. \blacksquare

Proof of Theorem 3:

First we note that because $G_0(\mu, \sigma)$ is a continuous probability measure, if $G(\mu, \sigma)$ is a homogenous NRM with Lévy intensity $\rho(dv)G_0(d\mu, d\sigma)$, then using the stick-breaking representation, we have that the two marginal distributions $G_\mu(\mu)$ and $G_\sigma(\sigma)$ are also homogenous NRMs with

Lévy intensities $\rho(dv)G_{0,\mu}(d\mu)$ and $\rho(dv)G_{0,\sigma}(\sigma)$ respectively. Given the conclusion of Theorem 2, we have that if $G_{0,\mu}$ or $G_{0,\sigma}$ satisfies (i) of Theorem 2, then $\alpha_+(G_\mu) = 0$ or $\alpha_+(G_\sigma) = 0$; if $G_{0,\mu}$ or $G_{0,\sigma}$ satisfies (ii) of Theorem 2, then $\alpha_+(G_\mu) = +\infty$ or $\alpha_+(G_\sigma) = +\infty$.

Since $k(\cdot)$ has part of the support in \mathbb{R}^+ , $E_K X_+^m > 0$ for any $m > 0$. We will examine the existence of moments $E_F X_+^m$ with F from Model 7 for any $m > 0$, and use Lemma A.2 to determine $\alpha_+(F)$. Similar to the proof of Theorem 1, we can analysis $E_F X_+^m$ using the lower bounds and the upper bound from Lemma A.4.

(i) If $\alpha_+(G_\mu) = 0$, then $E_{G_\mu} \mu_+^m = +\infty$ for all $m > 0$. Also note that $\overline{K}(0) > 0$ and $E_K X_+^m > 0$ since $k(\cdot)$ has full support in \mathbb{R} . Therefore, by the lower bound (A.2), $E_F X_+^m = +\infty$ for all $m > 0$ since $\overline{K}(0) > 0$, $E_K X_+^m > 0$, and $E_{G_{\mu,\sigma}} [\sigma^m I(\mu \geq 0)] \geq 0$. This implies $\alpha_+(F) = 0$ by Lemma A.2.

(ii) If $\alpha_+(G_\mu) = +\infty$ and $\alpha_+(G_\sigma) = 0$, then for all $m > 0$, $E_{G_\mu} \mu_+^m < +\infty$ and $E_{G_\sigma} \sigma^m = +\infty$. Because we have assumed that $G_{0,\mu}$ is symmetric about zero, this implies that $E_{G_\mu} \mu_-^m < +\infty$ for all $m > 0$. Also $E_K X_+^m > 0$ for all $m > 0$. Therefore by the lower bound (A.3), $E_F X_+^m = +\infty$. This again implies $\alpha_+(F) = 0$ by Lemma A.2.

(iii) If $\alpha_+(G_\mu) = +\infty$ and $\alpha_+(G_\sigma) = +\infty$, then for all $m > 0$, $E_{G_\mu} \mu_+^m < +\infty$ and $E_{G_\sigma} \sigma^m < +\infty$. Therefore, if $m \in (0, +\infty)$ and $m < \alpha_+(K)$, $E_K X_+^m < +\infty$ and $E_F X_+^m < +\infty$ by the upper bound (A.4); if $m \in (0, +\infty)$ and $m > \alpha_+(K)$, $E_K X_+^m = +\infty$ and $E_F X_+^m = +\infty$ by the lower bound (A.3). Similar analysis follows for the cases where $m = 0$ and $m = +\infty$ and we obtain that $\alpha_+(F) = \alpha_+(K)$.

We can summarize the results from different scenarios in the following table:

Table 1: $\alpha_+(F)$ in Theorem 3			
Values of Tail Indices: $a \in [0, +\infty]$			
$\alpha_+(G_\mu)$	$\alpha_+(G_\sigma)$	$\alpha_+(K)$	$\alpha_+(F)$
0	0	a	0
0	$+\infty$	a	0
$+\infty$	0	a	0
$+\infty$	$+\infty$	a	a

In every scenario listed in the table, regardless of the tail index $\alpha_+(K)$ of the kernel, $\alpha_+(F)$ is always a fixed number. Therefore, $\alpha_+(F)$ is almost surely a singleton, if both $G_{0,\mu}(\mu)$ and $G_{0,\sigma}(\sigma)$ satisfy either (i) or (ii) in Theorem 2. \blacksquare

Proof of Theorem 4:

We will show that

(a) If $\overline{G}_0(x) \succ \log \log x / \log x$, then part (i) of Theorem 2 holds;

(b) If $\overline{G}_0(x) \prec 1 / [(\log x) \cdot (\log \log x)^\delta]$ for some $\delta > 1$, then part (ii) of Theorem 2 holds.

If both $\overline{G}_{0,\mu}$ and $\overline{G}_{0,\sigma}$ satisfy either (a) or (b), then their right tail indices are either 0 or $+\infty$, and the conclusion of Theorem 4 follows directly from Theorem 3.

To show (a) and (b), we use the similar arguments as in [11]. We note that a cdf $G(x)$ drawn from $\text{DP}(a, G_0(x))$ can be written as a normalized Gamma process with Lévy intensity $\rho(dv)G_0(dx) = av^{-1}e^{-v}dvG_0(dx)$. The Laplace exponent for ρ is $\Psi(s) = a \log(1 + s)$ and its inversion is $\Psi^{-1}(u) = e^{u/a} - 1$. Thus for any given $\gamma > 1$, the function (9) in Proposition 1 is given by

$$h_\gamma(x) = \frac{\log |\log x|}{\exp\left(\frac{\gamma \log |\log x|}{ax} - 1\right)}$$

and $h_\gamma(x) \geq 0$ in $[0, \epsilon)$ for small enough $\epsilon > 0$. Using $\overline{G}_0(x) \succ \log \log x / \log x$ in (a), we have for all sufficiently large x (in other words sufficiently small $\overline{G}_0(x)$),

$$\begin{aligned} \frac{-\log h_\gamma(\overline{G}_0(x))}{\log x} &= \frac{\log \left[\exp\left(\frac{\gamma \log |\log \overline{G}_0(x)|}{a\overline{G}_0(x)}\right) - 1 \right] - \log \log |\log \overline{G}_0(x)|}{\log x} \\ &\leq \frac{\gamma \log |\log \overline{G}_0(x)|}{a\overline{G}_0(x) \log x} \prec \frac{\log \log \log x}{\log \log x} \rightarrow 0, \end{aligned}$$

which is exactly the condition in part (i) of Theorem 2. Thus (a) is proved. For (b), we set $h(x) = \exp \left[- \left(x |\log x|^{\delta'} \right)^{-1} \right]$ for some $1 < \delta' < \delta$. This function is convex in $[0, \epsilon)$ for small enough $\epsilon > 0$. Due to the lower and upper bounds $e^{-1} \log 1/u \leq \rho[u, +\infty) \leq \log(1/u) + e^{-1}$ (see [11]) and $\delta' > 1$, we have $\int_0^\epsilon \rho[h(x), +\infty) dx < +\infty$. Furthermore, if $\overline{G}_0(x) \prec 1/[(\log x) \cdot (\log \log x)^\delta]$, then for all sufficiently large x ,

$$\begin{aligned} \frac{-\log h(\overline{G}_0(x))}{\log x} &= \frac{1}{\overline{G}_0(x) |\log \overline{G}_0(x)|^{\delta'} \log x} \\ &\succ \frac{(\log x) \cdot (\log \log x)^\delta}{(\log \log x)^{\delta'} \log x} = (\log \log x)^{\delta - \delta'} \rightarrow +\infty, \end{aligned}$$

which is exactly the condition in part (ii) of Theorem 2. Thus (b) is proved. ■

Proof of Theorem 5:

We will show that

(a) If $\overline{G}_0(x) \succ x^{-\delta}$ for all $\delta > 0$, then part (i) of Theorem 2 holds;

(b) If $\overline{G}_0(x) \prec x^{-\delta}$ for all $\delta > 0$, then part (ii) of Theorem 2 holds.

If both $\overline{G}_{0,\mu}$ and $\overline{G}_{0,\sigma}$ satisfy either (a) or (b), then their right tail indices are either 0 or $+\infty$, and the conclusion of Theorem 5 follows directly from Theorem 3.

To show (a), we note that for the Lévy process with intensity $\rho(dv) = \frac{a}{\Gamma(1-\kappa)} v^{-\kappa-1} e^{-\tau v} dv$, its Laplace exponent is $\Psi(s) = \frac{a}{\kappa} [(s + \tau)^\kappa - \tau^\kappa]$, and its inverse is $\Psi^{-1}(u) = [\kappa u/a + \tau^\kappa]^{1/\kappa} - \tau$. Thus for any given $\gamma > 1$, the function (9) in Proposition 1 is given by

$$h_\gamma(x) = \frac{\log |\log x|}{\left[\frac{\kappa \log |\log x|}{ax} + \tau^\kappa \right]^{1/\kappa} - \tau}$$

and $h_\gamma(x) \geq 0$ in $[0, \epsilon)$ for small enough $\epsilon > 0$. For all sufficiently large x , we have

$$\frac{-\log h_\gamma(\overline{G}_0(x))}{\log x} \leq \frac{\frac{1}{\kappa} \log \frac{2\kappa \log |\log \overline{G}_0(x)|}{a\overline{G}_0(x)} - \log \log |\log \overline{G}_0(x)|}{\log x}$$

$$= \frac{-\frac{1}{\kappa} \log \overline{G}_0(x) + (\frac{1}{\kappa} - 1) \log \log |\log \overline{G}_0(x)| + \frac{1}{\kappa} \log \frac{2\kappa}{a}}{\log x} \preceq \frac{-\log \overline{G}_0(x)}{\log x} \rightarrow 0,$$

where the last relation follows from the assumption in (a) that for all $\delta > 0$, $\overline{G}_0(x) \succ x^{-\delta} \implies -\log \overline{G}_0(x) \prec \log x$. Thus (a) is proved.

For (b), we have the following bound for $\rho[u, +\infty)$:

$$\begin{aligned} \rho[u, +\infty) &= \frac{a}{\Gamma(1-\kappa)} \int_u^1 t^{-\kappa-1} e^{-\tau t} dt + \frac{a}{\Gamma(1-\kappa)} \int_1^\infty t^{-\kappa-1} e^{-\tau t} dt \\ &\leq \frac{a}{\Gamma(1-\kappa)} \int_u^1 t^{-\kappa-1} dt + \frac{a}{\Gamma(1-\kappa)} \int_1^\infty e^{-\tau t} dt \\ &= \frac{a}{\kappa \Gamma(1-\kappa)} \left(\frac{1}{u^\kappa} - 1 \right) + \frac{ae^{-\tau}}{\tau \Gamma(1-\kappa)}. \end{aligned}$$

Therefore if $\int_0^\epsilon \frac{1}{h(x)^\kappa} dx < +\infty$, then $\int_0^\epsilon \rho[h(x), +\infty) dx < +\infty$. We let $h(x) = (x|\log x|^\eta)^{1/\kappa}$ for some $\eta > 1$. For $\kappa \in (0, 1)$, this $h(x)$ is nonnegative and convex in $[0, \epsilon)$ if ϵ is sufficiently small. Furthermore, we have

$$\frac{-\log h(\overline{G}_0(x))}{\log x} = \frac{-\frac{1}{\kappa} \log \overline{G}_0(x) - \frac{\eta}{\kappa} \log |\log \overline{G}_0(x)|}{\log x} \rightarrow +\infty,$$

which follows from the assumption in part (b) that for all $\delta > 0$, $\overline{G}_0(x) \prec x^{-\delta} \implies -\log \overline{G}_0(x) \succ \log x$. Thus (b) is proved. \blacksquare

B Technical Proofs for Section 3

Proof of Theorem 7:

We show that for $\epsilon > 0$, the test $\Phi_n = I(|\hat{\alpha}_{s_n} - \alpha_{0+}| \geq \epsilon/2)$ with s_n in (iv) and $\hat{\alpha}_{s_n}$ satisfies (UT). Let $p_{s_n} = P_F(X > e^{s_n})$ be the population mean of \hat{p}_{k_n} and let $\alpha_{s_n} = \log(p_{s_n}) - \log(p_{s_n+1})$. Note that p_{s_n} and α_{s_n} implicitly depend on F . We complete the proof in two steps.

Step 1: Show $E_{F_0} \Phi_n \rightarrow 0$ as $n \rightarrow \infty$.

We have

$$\begin{aligned} E_{F_0} \Phi_n &= P_{F_0} \left(|\hat{\alpha}_{s_n} - \alpha_{0+}| \geq \frac{\epsilon}{2} \right) \\ &\leq P_{F_0} \left(|\hat{\alpha}_{s_n} - \alpha_{s_n}| \geq \frac{\epsilon}{4} \right) + P_{F_0} \left(|\alpha_{s_n} - \alpha_{0+}| \geq \frac{\epsilon}{4} \right). \end{aligned} \quad (\text{A.7})$$

The first term in (A.7) can be bounded by Lemma 2 and equation (4.2) of [5]:

$$P_{F_0} \left(|\hat{\alpha}_{s_n} - \alpha_{s_n}| \geq \frac{\epsilon}{4} \right) \leq 2 \exp \left(-\frac{np_{s_n+1}\epsilon^2}{144} \right), \quad (\text{A.8})$$

where $p_{s_n+1} = P_{F_0}(X > e^{s_n+1}) = e^{-\alpha_{0+}(s_n+1)} L_0(e^{s_n+1})$. Since L_0 is slowly varying, as $n \rightarrow \infty$, eventually $L_0(e^{s_n+1}) \geq e^{-\delta(s_n+1)}$ for arbitrarily small $\delta > 0$. By Condition (iv) of Theorem 7, $s_n \prec \log n/\alpha_{0+}$ and hence $p_{s_n+1} \geq \exp(-(\alpha_{0+} + \delta)(s_n + 1)) \rightarrow +\infty$, which implies that the righthand side of (A.8) goes to zero as $n \rightarrow \infty$.

The second term in (A.7) is not stochastic. We have

$$|\alpha_{s_n} - \alpha_{0+}| = |\log(p_{s_n}) - \log(p_{s_n+1}) - \alpha_{0+}|$$

$$= \left| \log \frac{L_0(e^{s_n})}{L_0(e^{s_n+1})} \right| \rightarrow 0,$$

because L_0 is slowly varying and $s_n \rightarrow \infty$. Therefore both terms on the righthand side of (A.7) converge to zero as $n \rightarrow \infty$.

Step 2: Show $\sup_{F \in B_{\alpha+}^c(F_0, \epsilon) \cap \mathcal{F}_n} E_F(1 - \Phi_n) \rightarrow 0$ as $n \rightarrow \infty$, where we let $\mathcal{F}_n = \mathcal{F}_{1n} \cap \mathcal{F}_{2n} \cap \mathcal{F}_{3n}$. By Conditions (i)-(iii), it is clear that $\Pi(\mathcal{F}_n^c) \leq \Pi(\mathcal{F}_{1n}^c) + \Pi(\mathcal{F}_{2n}^c) + \Pi(\mathcal{F}_{3n}^c) \leq e^{-c_1 n} + e^{-c_2 n} + e^{-c_3 n} \leq e^{-c' n}$ where $c' = \min(c_1, c_2, c_3)/2$.

For every $F \in B_{\alpha+}^c(F_0, \epsilon) \cap \mathcal{F}_n$, we have $|\alpha_+(F) - \alpha_{0+}| > \epsilon$. Therefore

$$\begin{aligned} E_F(1 - \Phi_n) &= P_F \left(|\hat{\alpha}_{s_n} - \alpha_{0+}| \leq \frac{\epsilon}{2} \right) \\ &= P_F \left(|\hat{\alpha}_{s_n} - \alpha_{0+}| \leq \frac{\epsilon}{2}, |\hat{\alpha}_{s_n} - \alpha_+(F)| < \frac{\epsilon}{2} \right) + P_F \left(|\hat{\alpha}_{s_n} - \alpha_{0+}| \leq \frac{\epsilon}{2}, |\hat{\alpha}_{s_n} - \alpha_+(F)| \geq \frac{\epsilon}{2} \right) \\ &\leq P_F(|\alpha_+(F) - \alpha_{0+}| < \epsilon) + P_F \left(|\hat{\alpha}_{s_n} - \alpha_+(F)| \geq \frac{\epsilon}{2} \right) \\ &= P_F \left(|\hat{\alpha}_{s_n} - \alpha_+(F)| \geq \frac{\epsilon}{2} \right) \leq P_F \left(|\hat{\alpha}_{s_n} - \alpha_{s_n}| \geq \frac{\epsilon}{4} \right) + P_F \left(|\alpha_{s_n} - \alpha_+(F)| \geq \frac{\epsilon}{4} \right). \end{aligned} \quad (\text{A.9})$$

We only need to show that both terms on the righthand side of (A.9) converge to zero uniformly over all $F \in B_{\alpha+}^c(F_0, \epsilon) \cap \mathcal{F}_n$ as $n \rightarrow \infty$. For a fixed F , the first term can be bounded by Lemma 2 and equation (4.2) of [5] again as

$$P_F \left(|\hat{\alpha}_{s_n} - \alpha_{s_n}| \geq \frac{\epsilon}{4} \right) \leq 2 \exp \left(-\frac{np_{s_n+1}\epsilon^2}{576} \right).$$

To obtain uniform convergence for the righthand side, we only need the quantity np_{s_n+1} to be uniformly bounded below for all $F \in B_{\alpha+}^c(F_0, \epsilon) \cap \mathcal{F}_n$. Using Conditions (ii) and (iii), we can obtain the following uniform lower bound:

$$\begin{aligned} np_{s_n+1} &= ne^{-\alpha_+(F)(s_n+1)} L_F(e^{s_n+1}) = ne^{-\alpha_+(F)(s_n+1)} L_F(x_0) \exp \left(\int_{x_0}^x \frac{h_F(t)}{t} dt \right) \\ &\geq \exp \left(\log n - \bar{\alpha}_n(s_n+1) - c_0 - \int_{x_0}^x \frac{B_n}{t^{\tau_n+1}} dt \right) \\ &= \exp \left(\log n - \bar{\alpha}_n(s_n+1) - c_0 - \frac{B_n}{\tau_n x_0^{\tau_n}} + \frac{B_n}{\tau_n} e^{-\tau_n(s_n+1)} \right) \\ &\geq \exp \left(\log n - \bar{\alpha}_n(s_n+1) - c_0 - \frac{B_n}{\tau_n} \right). \end{aligned}$$

From Condition (iv), $\log n \succ \bar{\alpha}_n(s_n+1)$ and $\log n \succ \frac{B_n}{\tau_n}$. Therefore we have obtained that uniformly over all $F \in B_{\alpha+}^c(F_0, \epsilon) \cap \mathcal{F}_n$, $P_F(|\hat{\alpha}_{s_n} - \alpha_{s_n}| \geq \frac{\epsilon}{4})$ converges to zero as $n \rightarrow \infty$.

For the second term in (A.9), we have

$$\begin{aligned} |\alpha_{s_n} - \alpha_+(F)| &= |\log p_{s_n} - \log p_{s_n+1} - \alpha_+(F)| \\ &= \left| \log e^{-\alpha_+(F)s_n} L_F(e^{s_n}) - \log e^{-\alpha_+(F)(s_n+1)} L_F(e^{s_n+1}) - \alpha_+(F) \right| \\ &= |\log L_F(e^{s_n}) - \log L_F(e^{s_n+1})| = \left| \int_{e^{s_n}}^{e^{s_n+1}} \frac{h_F(x)}{x} dx \right| \\ &\leq \left| \int_{e^{s_n}}^{e^{s_n+1}} \frac{\bar{h}_n(x)}{x} dx \right| \leq \sup_{x \in [e^{s_n}, e^{s_n+1}]} |\bar{h}_n(x)| = B_n e^{-\tau_n s_n}, \end{aligned}$$

and $B_n e^{-\tau_n s_n} \rightarrow 0$ by Condition (iv). Therefore the probability $P_F(|\alpha_{s_n} - \alpha_+(F)| \geq \frac{\epsilon}{4})$ is zero for all large n uniformly over all $F \in B_{\alpha_+}^c(F_0, \epsilon) \cap \mathcal{F}_n$. \blacksquare

Lemma A.5. *A distribution $F(x) \in \mathcal{CM}_e \cap \mathcal{P}_2$ if and only if the survival function $\overline{F}(x)$ and the density $f(x)$ take the form*

$$\begin{aligned}\overline{F}(x) &= wx^{-\alpha} + (1-w) \int_{\alpha(1+\beta)}^{+\infty} x^{-u} dH(u), \\ f(x) &= w\alpha x^{-(\alpha+1)} + (1-w) \int_{\alpha(1+\beta)}^{+\infty} ux^{-(u+1)} dH(u),\end{aligned}$$

where $w \in (0, 1]$, $\alpha > 0$, $\beta > 0$, $H(\alpha)$ is a probability measure whose support is in $[\alpha(1+\beta), +\infty)$.

Proof of Lemma A.5:

We only need to show the expression of $\overline{F}(x)$ since the expression of $f(x)$ will follow directly. We note that if $F(x)$ has a density $f(x)$, then $F(e^t)$ for $t \in [0, +\infty)$ is also a cdf and $\overline{F}(e^t)$ is a survival function. By the Hausdorff-Bernstein-Widder theorem, $\overline{F}(e^t)$ is a completely monotone function on $t \in [0, +\infty)$ if and only if it is the Laplace transformation of some probability distribution $G(\alpha)$ on $\alpha \in (0, +\infty)$, i.e.

$$\overline{F}(e^t) = \int_0^\infty e^{-ut} dG(u),$$

which is equivalent to say that for $x \in [1, +\infty)$,

$$\overline{F}(x) = \int_1^\infty x^{-u} dG(u). \quad (\text{A.10})$$

Therefore to prove the conclusion of the lemma, it only remains to show that $F(x) \in \mathcal{P}_2$ if and only if the probability measure G has the decomposition

$$G(u) = w\delta_\alpha + (1-w)H(u), \quad (\text{A.11})$$

for some $w \in (0, 1]$, $\alpha > 0$ and some probability measure $H(u)$ supported on $[\alpha(1+\beta), +\infty)$.

It is clear that if G has the form (A.11) (so that $\overline{F}(x)$ has the form in the lemma), then $F(x) \in \mathcal{P}_2$. Conversely, if $F(x)$ has the form of (A.10) and meanwhile $F(x) = Cx^{-\alpha} + O(x^{-\alpha(1+\beta)})$ for some $\alpha, \beta > 0$ and $C > 0$, then we show that:

- (i) $G(0, \alpha) = 0$;
- (ii) $G(\alpha, \alpha(1+\beta)) = 0$.

If (i) does not hold, then there exists a set $I \subset (0, \alpha)$ such that $G(I) > 0$. Thus $\int_I x^{-u} dG(u) \asymp x^{-\alpha}$ which contradicts the fact that $\limsup_{x \rightarrow +\infty} x^\alpha \overline{F}(x) = C < +\infty$. For (ii), we have

$$\limsup_{x \rightarrow +\infty} x^{\alpha(1+\beta)} \int_0^\infty x^{-u} d[G(u) - C\delta_\alpha(u)] < +\infty.$$

Therefore by similar argument to the proof of part (i), the signed measure $G - C\delta_\alpha$ on $[\alpha, +\infty)$ has no support in $(\alpha, \alpha(1+\beta))$, which is equivalent to part (ii). (i) and (ii) together imply that

$G - C\delta_\alpha$ is a signed measure supported on $[\alpha(1 + \beta), +\infty)$. Because G is a probability measure, it must hold that $w = C \in (0, 1]$ and $(G - w\delta_\alpha)/(1 - w)$ is a probability measure. Thus we have proved that $F(x) \in \mathcal{CM}_e \cap \mathcal{P}_2$ implies (A.11). \blacksquare

Proof of Theorem 9:

We prove the theorem in a similar way to the proof of the exponential mixture model in Theorem 16 of [47].

By Lemma A.5, we assume that the true density function has the form

$$f_0(x) = w_0 x^{-\alpha_0} + \int_{\alpha_0(1+\beta_0)}^{\infty} \alpha x^{-(\alpha+1)} dG_0(\alpha), \quad (\text{A.12})$$

where α_0 is short for $\alpha_{0+}(F)$ and G_0 is supported on $[\alpha_0(1 + \beta_0), +\infty)$. Without causing confusion, we also denote a generic $f(x)$ from $\mathcal{CM}_e \cap \mathcal{P}_2$ by $f_{w_1, \alpha_1, H}(x)$ which takes the form of Model (12).

We use Π to denote the prior measure. The KL condition for Model (12) is equivalent to say that for any $\epsilon > 0$, there exists sets $\mathcal{W} \subset [\underline{w}_n, 1]$ (for w_1), $\mathcal{A} \subset (0, +\infty)$ (for α_1) and \mathcal{H} (for H), and $\liminf_{n \rightarrow \infty} \Pi(\mathcal{W} \times \mathcal{A} \times \mathcal{H}) > 0$, such that for all $(w_1, \alpha_1, H) \in \mathcal{W} \times \mathcal{A} \times \mathcal{H}$ and all sufficiently large n ,

$$\int_1^{+\infty} f_0(x) \log \frac{f_0(x)}{f_{w_1, \alpha_1, H}(x)} dx < \epsilon.$$

By Theorem 1 of [47], we prove the following 3 relations to obtain the conclusion for all $\epsilon \in (0, 1/2)$ and all sufficiently large n :

- (i) $\int_1^{\infty} f_0(x) \log \frac{f_0(x)}{f_{w_0, \alpha_0, G_1}(x)} dx < \frac{\epsilon}{3}$ for some (ϵ -dependent) G_1 supported on $[\alpha_0(1 + \beta_0), \bar{\alpha}_n + \tau_n]$;
- (ii) $\int_1^{\infty} f_0(x) \log \frac{f_{w_0, \alpha_0, G_1}(x)}{f_{w_1, \alpha_1, G_1}(x)} dx < \frac{\epsilon}{3}$ for $(w_1, \alpha_1) \in \mathcal{W} \times \mathcal{A}$ and $\Pi(\mathcal{W} \times \mathcal{A}) > 0$;
- (iii) $\int_1^{\infty} f_0(x) \log \frac{f_{w_1, \alpha_1, G_1}(x)}{f_{w_1, \alpha_1, H}(x)} dx < \frac{\epsilon}{3}$ for all $H \in \mathcal{H}$, all $(w_1, \alpha_1) \in \mathcal{W} \times \mathcal{A}$ and $\Pi(\mathcal{H}) > 0$.

Check (i): For $a > 0$, we have that for all large n , $a \leq \bar{\alpha}_n + \tau_n$, and $[\alpha_0(1 + \beta_0), a] \subset [\alpha_0(1 + \beta_0), \bar{\alpha}_n + \tau_n]$. Let $G_1(A) = G_0(A)/G_0([\alpha_0(1 + \beta_0), a])$ for $A \subseteq [\alpha_0(1 + \beta_0), +\infty)$. Then as $a \rightarrow +\infty$ and $n \rightarrow +\infty$, for every fixed $x \geq 1$, $f_{w_0, \alpha_0, G_1}(x) \rightarrow f_{w_0, \alpha_0, G_0}(x) = f_0(x)$. Therefore, for sufficiently large a and n , $f_{w_0, \alpha_0, G_1}(x) \leq 2f_0(x)$. This is an upper bound for $f_{w_0, \alpha_0, G_1}(x)$. To get a lower bound, we notice that for fixed $x > 1$, the function $\alpha x^{-(\alpha+1)}$ attains its maximum when $\alpha = 1/\log x$, increases on $(0, 1/\log x]$ and decreases on $(1/\log x, +\infty)$. For sufficiently large a , we pick a fixed $\alpha_2 \in (\alpha_0(1 + \beta_0), a)$ such that $G_0[\alpha_0(1 + \beta_0), \alpha_2] > 1/2$. Then the following relation holds:

$$\begin{aligned} \int_{\alpha_0(1+\beta_0)}^a u x^{-(u+1)} dG_1(u) &\geq \frac{1}{2} \alpha_2 x^{-(\alpha_2+1)} \quad \text{if } x \geq e^{1/[\alpha_0(1+\beta_0)]}, \\ \int_{\alpha_0(1+\beta_0)}^a u x^{-(u+1)} dG_1(u) &\geq \frac{1}{2} \min \left(\alpha_0(1 + \beta_0) x^{-(\alpha_0(1+\beta_0)+1)}, \alpha_2 x^{-(\alpha_2+1)} \right) \\ &\quad \text{if } e^{1/\alpha_2} \leq x < e^{1/[\alpha_0(1+\beta_0)]}, \\ \int_{\alpha_0(1+\beta_0)}^a u x^{-(u+1)} dG_1(u) &\geq \frac{1}{2} \alpha_0(1 + \beta_0) x^{-(\alpha_0(1+\beta_0)+1)} \quad \text{if } 1 \leq x < e^{1/\alpha_2}. \end{aligned}$$

This gives a lower bound for $f_{w_0, \alpha_0, G_1}(x)$:

$$f_{w_0, \alpha_0, G_1}(x) \geq w_0 \alpha_0 x^{-(\alpha_0+1)} + \frac{1}{2} \min \left(\alpha_0(1 + \beta_0) x^{-(\alpha_0(1+\beta_0)+1)}, \alpha_2 x^{-(\alpha_2+1)} \right).$$

Therefore, we can combine the upper and lower bounds and obtain that

$$|\log f_{w_0, \alpha_0, G_1}(x)| \leq \max \left\{ \log 2 + |\log f_0(x)|, \log 2 + |\log \alpha_0(1 + \beta_0)| + [\alpha_0(1 + \beta_0) + 1] \log x, \right. \\ \left. \log 2 + |\log \alpha_2| + (\alpha_2 + 1) \log x \right\}.$$

Since clearly both $\log x$ and $|\log f_0(x)|$ are f_0 -integrable, we obtain from above that $|\log f_{w_0, \alpha_0, G_1}(x)|$ is also f_0 -integrable. Therefore by the dominated convergence theorem, as $a \rightarrow +\infty$ and $n \rightarrow +\infty$, $\int_1^\infty f_0(x) \log \frac{f_0(x)}{f_{w_0, \alpha_0, G_1}(x)} dx \rightarrow 0$. Hence (i) is proved.

Check (ii): We show (ii) for a fixed G_1 that satisfies part (i). We have also shown that $|\log f_{w_0, \alpha_0, G_1}(x)|$ is f_0 -integrable. Let $\mathcal{W} = [w_0 - \eta, w_0 + \eta]$ and $\mathcal{A} = [\alpha_0 - \eta, \alpha_0 + \eta]$ for some $\eta > 0$. Then since $f_{w_1, \alpha_1, G_1}(x)$ is a continuous function of (w_1, α_1) at (w_0, α_0) , $f_{w_1, \alpha_1, G_1}(x) \rightarrow f_{w_0, \alpha_0, G_1}(x)$ pointwise in x , uniformly for all $w_1 \in \mathcal{W}$ and $\alpha_1 \in \mathcal{A}$ as $\eta \rightarrow 0$. Hence there exists $\eta > 0$ such that $f_{w_0, \alpha_0, G_1}(x)/2 \leq f_{w_1, \alpha_1, G_1}(x) \leq 2f_{w_0, \alpha_0, G_1}(x)$ for all $w_1 \in \mathcal{W}$ and $\alpha_1 \in \mathcal{A}$, which means $|\log f_{w_1, \alpha_1, G_1}(x)|$ is also f_0 -integrable. Therefore by the dominated convergence theorem, as $\eta \rightarrow 0$, $\int_1^\infty f_0(x) \log \frac{f_{w_0, \alpha_0, G_1}(x)}{f_{w_1, \alpha_1, G_1}(x)} dx \rightarrow 0$ and (ii) is proved.

Check (iii): The argument is similar to the proof of Lemma 3 in [47]. We split the integral into two parts:

$$\int_1^\infty f_0(x) \log \frac{f_{w_1, \alpha_1, G_1}(x)}{f_{w_1, \alpha_1, H}(x)} dx \leq \int_1^{C_1} f_0(x) \log \frac{f_{w_1, \alpha_1, G_1}(x)}{f_{w_1, \alpha_1, H}(x)} dx + \int_{C_1}^\infty f_0(x) \log \frac{f_{w_1, \alpha_1, G_1}(x)}{f_{w_1, \alpha_1, H}(x)} dx \\ := I_1 + I_2. \quad (\text{A.13})$$

Let $D = [\alpha_0 - \eta, a]$. Then $G_1(D) = 1$ and $\mathcal{H}_1 = \{H \sim \Pi : H(D) > 1/2\}$ is an open neighborhood of G_1 . For all $H \in \mathcal{H}_1$ and $(w_1, \alpha_1) \in \mathcal{W} \times \mathcal{A}$,

$$I_2 \leq \int_{C_1}^\infty f_0(x) \log \frac{f_{w_1, \alpha_1, G_1}(x)}{w_1 \alpha_1 x^{-(\alpha_1+1)} + (1 - w_1) \int_{\alpha_1 + \tau_n}^\infty \inf_u k(x; u) dH(u)} dx \\ \leq \int_{C_1}^\infty f_0(x) \log \frac{f_{w_1, \alpha_1, G_1}(x)}{w_1 \alpha_1 x^{-(\alpha_1+1)} + \frac{1}{2} (1 - w_1) \min(\alpha_1 x^{-(\alpha_1+1)}, \alpha_2 x^{-(\alpha_2+1)})} dx, \quad (\text{A.14})$$

where the 2nd inequality follows from $\inf_{u \geq \alpha_1 + \tau_n} k(x; u) \geq \min(\alpha_1 x^{-(\alpha_1+1)}, \alpha_2 x^{-(\alpha_2+1)})$ similar to the proof of part (i). It is clear that both $|\log f_{w_1, \alpha_1, G_1}(x)|$ and $|\log [w_1 \alpha_1 x^{-(\alpha_1+1)} + \frac{1}{2} (1 - w_1) \min(\alpha_1 x^{-(\alpha_1+1)}, \alpha_2 x^{-(\alpha_2+1)})]|$ are f_0 -integrable uniformly for all $(w_1, \alpha_1) \in \mathcal{W} \times \mathcal{A}$. Therefore we can choose C_1 sufficiently large, such that $I_2 < \epsilon/6$.

To bound I_1 , note that

$$I_1 \leq \sup_{x \in [1, C_1]} \left| \frac{f_{w_1, \alpha_1, G_1}(x)}{f_{w_1, \alpha_1, H}(x)} - 1 \right| \leq \sup_{x \in [1, C_1]} \left| \frac{\int_{\alpha_1 + \tau_n}^{+\infty} k(x; u) d[G_1(u) - H(u)]}{f_{w_1, \alpha_1, H}(x)} \right| \\ \leq \frac{\sup_{x \in [1, C_1]} \left| \int_{\alpha_1 + \tau_n}^{+\infty} k(x; u) d[G_1(u) - H(u)] \right|}{\inf_{x \in [1, C_1]} f_{w_1, \alpha_1, H}(x)}. \quad (\text{A.15})$$

For any measure H , all $(w_1, \alpha_1) \in \mathcal{W} \times \mathcal{A}$,

$$\inf_{x \in [1, C_1]} f_{w_1, \alpha_1, H}(x) \geq \inf_{x \in [1, C_1]} w_1 \alpha_1 C_1^{-(\alpha_1+1)} \geq (w_0 - \eta)(\alpha_0 - \eta) C_1^{-(\alpha_0 + \eta + 1)} := C_2. \quad (\text{A.16})$$

Within \mathcal{H}_1 , we further choose an open weak neighborhood \mathcal{H}_2 of G_1 such that $H[\alpha_0 - \eta, a] > 1 - C_2\epsilon/(24a)$ if $H \in \mathcal{H}_2$. Then for all $H \in \mathcal{H}_2$,

$$\sup_{x \in [1, C_1]} \left| \int_{D^c} k(x; u) d[G_1(u) - H(u)] \right| \leq a(G_1(D^c) + H(D^c)) < \frac{C_2\epsilon}{24} \quad (\text{A.17})$$

since $G_1(D^c) = 0$. Because the kernel $k(x; \alpha)$ is equicontinuous on D , by the Arzela-Ascoli theorem, there exist N points $x_1, \dots, x_N \in [1, C_1]$ such that for any $x \in [1, C_1]$, $\sup_{u \in D} |k(x; u) - k(x_i; u)| < C_2\epsilon/24$ for some $i = 1, \dots, N$. Now choose a smaller open weak neighborhood $\mathcal{H}_3 \subseteq \mathcal{H}_2$ such that $\max_{i=1, \dots, N} \left| \int_D k(x_i; u) d[G_1(u) - H(u)] \right| < C_2\epsilon/24$ for all $H \in \mathcal{H}_3$. Then for any $x \in [1, C_1]$ and $H \in \mathcal{H}_3$, there exists some x_i ($i = 1, \dots, N$) such that

$$\begin{aligned} & \left| \int_{\alpha_1 + \tau_n}^{+\infty} k(x; u) d[G_1(u) - H(u)] \right| \\ & \leq \left| \int_D k(x; u) d[G_1(u) - H(u)] \right| + \left| \int_{D^c} k(x; u) d[G_1(u) - H(u)] \right| \\ & \leq \left| \int_D [k(x; u) - k(x_i; u)] dG_1(u) \right| + \left| \int_D [k(x; u) - k(x_i; u)] dH(u) \right| \\ & \quad + \left| \int_D k(x_i; u) d[G_1(u) - H(u)] \right| + \left| \int_{D^c} k(x; u) d[G_1(u) - H(u)] \right| \\ & < \frac{C_2\epsilon}{8} + \left| \int_{D^c} k(x; u) d[G_1(u) - H(u)] \right|. \end{aligned} \quad (\text{A.18})$$

For $\epsilon \in (0, 1/2)$, (A.17) and (A.18) together give us

$$\sup_{x \in [1, C_1]} \left| \int_{\alpha_1 + \tau_n}^{+\infty} k(x; u) d[G_1(u) - H(u)] \right| \leq \frac{C_2\epsilon}{6} < \frac{1}{2}, \quad (\text{A.19})$$

for small $\epsilon > 0$ and all $H \in \mathcal{H}_3$. We combine (A.15), (A.16) and (A.19) and obtain that $I_1 < \frac{C_2\epsilon}{6C_2} = \frac{\epsilon}{6}$. Let $\mathcal{H} = \mathcal{H}_3$ and $\Pi(\mathcal{H}) > 0$ since it is an open neighborhood of G_1 . Therefore we have shown that in (A.13), for all $(w_1, \alpha_1) \in \mathcal{W} \times \mathcal{A}$ and all $H \in \mathcal{H}_3$,

$$\int_1^\infty f_0(x) \log \frac{f_{w_1, \alpha_1, G_1}(x)}{f_{w_1, \alpha_1, H}(x)} dx \leq I_1 + I_2 < \frac{\epsilon}{3},$$

which is the conclusion of part (iii). ■

Proof of Theorem 10:

Condition (i) implies (KL) by Theorem 9. We only need to show (UT) by checking the conditions of Theorem 7. For a distribution $F(x)$ drawn from Model (12), the slowly varying function $L_F(x)$ can be written as

$$L_F(x) = w_1 + (1 - w_1) \int_{\alpha_1 + \tau_n}^{+\infty} x^{\alpha_1 - \alpha} dH(\alpha).$$

Therefore, $L_F(1) = 1$. We can set $x_0 = 1$, $c_0 = 0.1$, and Condition (i) of Theorem 7 holds. Because $\alpha_+(F) = \alpha_1$ and α_1 is drawn from G_1 with support $(0, \bar{\alpha}_n]$, Condition (iii) of Theorem 7 is satisfied by the same assumption on $\bar{\alpha}_n$ in Condition (ii) of Theorem 10.

Finally we check the uniform bound on the function $h_F(x)$. In this case, $h_F(x)$ can be bounded as

$$|h_F(x)| = \left| \frac{xL'_F(x)}{L_F(x)} \right| = \left| \frac{(1 - w_1) \int_{\alpha_1 + \tau_n}^{+\infty} (\alpha_1 - \alpha) x^{\alpha_1 - \alpha} dH(\alpha)}{w_1 + \int_{\alpha_1 + \tau_n}^{+\infty} x^{\alpha_1 - \alpha} dH(\alpha)} \right| \leq \frac{\tau_n + E_{\tilde{H}}(\alpha)}{w_1} x^{-\tau_n},$$

where \tilde{H} is H shifted by $\alpha_1 + \tau_n$ to the left. Under Model (12), \tilde{H} can also be viewed as a draw from MFM or NRM with base measure H_0 . Since the support of H_0 is upper bounded by $\bar{\alpha}_n$, so is the support of \tilde{H} . Therefore $E_{\tilde{H}}\alpha \leq \bar{\alpha}_n$. Thus by $\tau_n \prec 1$ we can set $B_n = \frac{2\bar{\alpha}_n}{\underline{w}_n}$ and it follows that $|h_F(x)| \leq B_n x^{-\tau_n}$ uniformly for all F generated from Model (12), where τ_n is the same τ_n here in Theorem 10. By Condition (ii) of Theorem 10, $B_n = 2\bar{\alpha}_n/\underline{w}_n \prec \tau_n \log n$. Moreover, because $\log(\bar{\alpha}_n/\underline{w}_n) \prec \tau_n \log n / \bar{\alpha}_n$, there exists some sequence s_n such that $1 \prec s_n \prec \log n / \bar{\alpha}_n$ and $\log(\bar{\alpha}_n/\underline{w}_n) \prec \tau_n s_n$. For example,

$$s_n = \left[\frac{(\log n) (\log(\bar{\alpha}_n/\underline{w}_n))}{\bar{\alpha}_n \tau_n} \right]^{1/2}.$$

Hence $B_n = 2\bar{\alpha}_n/\underline{w}_n \prec \exp(\tau_n s_n)$. This has proved Conditions (ii) and (iv) of Theorem 7. ■

References

- [1] E. Barrios, A. Lijoi, L. E. Nieto-Barajas, and I. Prünster. Modeling with normalized random measure mixture models. *Statistical Science*, 28:313 – 334, 2013.
- [2] J. Beirlant, Y. Goegebeur, J. Teugels, J. Segers, D. De Waal, and C. Ferro. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons Ltd, 2004.
- [3] L. Bottolo, G. Consonni, P. Dellaportas, and A. Lijoi. Bayesian analysis of extreme values by mixture modeling. *Extremes*, 6:25 – 47, 2002.
- [4] S. Boucheron and M. Thomas. Tail index estimation, concentration and adaptivity. *arXiv: 1503.05077.*, 2015.
- [5] A. Carpetier and A. K. H. Kim. Adaptive and minimax optimal estimation of the tail coefficient. *Statistica Sinica*, 25:1133 – 1144, 2015.
- [6] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51:661 – 703, 2009.

- [7] U. Cormann and R.-D. Reiss. Generalizing the Pareto to the log-Pareto model and statistical inference. *Extremes*, 12:93 – 105, 2009.
- [8] L. De Haan and S. I. Resnick. A simple asymptotic estimate for the index of a stable distribution. *Journal of the Royal Statistical Society, Series B*, 42:83 – 87, 1980.
- [9] J. Diebolt, M. A. El-Aroui, M. Garrido, and S. Girard. Quasi-conjugate Bayes estimates for GPD parameters and application to heavy tails modelling. *Extremes*, 8:57 – 78, 2005.
- [10] F. F. do Nascimento, D. Gamerman, and H. F. Lopes. A semiparametric Bayesian approach to extreme value estimation. *Statistics and Computing*, 22:661 – 675, 2012.
- [11] H. Doss and T. Sellke. The tails of probabilities chosen from a Dirichlet prior. *The Annals of Statistics*, 10:1302 – 1305, 1982.
- [12] H. Drees. Optimal rates of convergence for estimates of the extreme value index. *The Annals of Statistics*, 26:434 – 448, 1998.
- [13] H. Drees. Minimax risk bounds in extreme value theory. *The Annals of Statistics*, 29:266 – 294, 2001.
- [14] M. D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90:577 – 588, 1995.
- [15] S. Favaro and Y. W. Teh. MCMC for normalized random measures mixture models. *Statistical Science*, 28:335 – 359, 2013.
- [16] T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209 – 230, 1973.
- [17] A. Frigessi, O. Haug, and H. Rue. A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5:219 – 235, 2002.
- [18] B. E. Fristedt. Sample function behavior of increasing processes with stationary independent increments. *Pacific Journal of Mathematics*, 21:21 – 33, 1967.
- [19] B. E. Fristedt and W. E. Pruitt. Lower functions for increasing random walks and subordinators. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 18:167 – 182, 1971.
- [20] J. A. Fúquene P. A semi-parametric Bayesian extreme value model using a Dirichlet process mixture of gamma densities. *Journal of Applied Statistics*, 42:267 – 280, 2015.
- [21] S. Ghosal and A. W. van der Vaart. Posterior convergence rates of Dirichlet mixtures at smooth densities. *The Annals of Statistics*, 35:697 – 723, 2007.
- [22] S. Ghosal, J. K. Ghosh, and R. V. Ramamoorthi. Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics*, 27:143 – 158, 1999.

- [23] S. Ghosal, J. K. Ghosh, and A. W. van der Vaart. Convergence rates of posterior distributions. *The Annals of Statistics*, 28:500 – 531, 2000.
- [24] P. J. Green and S. Richardson. Modeling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics*, 28:355 – 375, 2001.
- [25] E. Haeusler and J. L. Teugels. On asymptotic normality of Hill’s estimator for the exponent of regular variation. *The Annals of Statistics*, 13:743 – 756, 1985.
- [26] P. Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society, Series B*, 44:37 – 42, 1982.
- [27] P. Hall and A. H. Welsh. Best attainable rates of convergence for estimates of parameters of regular variation. *The Annals of Statistics*, 12:1079 – 1084, 1984.
- [28] P. Hall and A. H. Welsh. Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13:331 – 341, 1984.
- [29] B. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163 – 1174, 1975.
- [30] L. F. James, A. Lijoi, and I. Prünster. Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistical*, 36:76 – 97, 2009.
- [31] W. Kruijer and J. Rousseau. Adaptive bayesian density estimation with location-scale mixtures. *Electronic Journal of Statistics*, 4:1225 – 1257, 2010.
- [32] A. Y. Lo. On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, 12:351 – 357, 1984.
- [33] S. N. MacEachern. Estimating normal means with a conjugate style Dirichlet process prior. *Communication in Statistics - Simulation and Computation*, 23:727 – 741, 1994.
- [34] D. M. Mason. Laws of large numbers for sums of extreme values. *The Annals of Probability*, 10:754 – 764, 1982.
- [35] R. M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249 – 265, 2000.
- [36] S. Y. Novak. Lower bounds to the accuracy of inference on heavy tails. *Bernoulli*, 20:979 – 989, 2014.
- [37] J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119 – 131, 1975.
- [38] E. Regazzini, A. Lijoi, and I. Prünster. Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics*, 31:560 – 585, 2003.

- [39] S. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B*, 59:731 – 792, 1997.
- [40] L. Schwartz. On Bayes procedures. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 4:10 – 26, 1965.
- [41] W. Shen, S. T. Tokdar, and S. Ghosal. Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. *Biometrika*, 100:623 – 640, 2013.
- [42] A. Stephenson and J. Tawn. Bayesian inference for extremes: accounting for the three extremal types. *Extremes*, 7:297 – 307, 2004.
- [43] S. T. Tokdar. Posterior consistency of Dirichlet location-scale mixture of normals in density estimation and regression. *Sankhyā*, 68:90 – 110, 2006.
- [44] J. Tressou. Bayesian nonparametrics for heavy tailed distribution. Application to food risk assessment. *Bayesian Analysis*, 3:367 – 392, 2008.
- [45] Z. Wang, A. Rodriguez, and A. Kottas. A nonparametric mixture modeling framework for extreme value analysis. Technical report, 2012.
<https://www.soe.ucsc.edu/research/technical-reports/UCSC-SOE-11-26/download>.
- [46] T. Watanabe. A probabilistic method in Hausdorff moment problem and the Laplace-Stieltjes transform. *Journal of the Mathematical Society of Japan*, 12:192 – 206, 1960.
- [47] Y. Wu and S. Ghosal. Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electronic Journal of Statistics*, 2:298 – 331, 2008.